



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, A61K 38/17, C12Q 1/68		A2	(11) International Publication Number: WO 00/15799 (43) International Publication Date: 23 March 2000 (23.03.00)																																								
(21) International Application Number: PCT/US99/21688 (22) International Filing Date: 17 September 1999 (17.09.99) (30) Priority Data: <table> <tr><td>60/155,246</td><td>17 September 1998 (17.09.98)</td><td>US</td></tr> <tr><td>09/158,720</td><td>22 September 1998 (22.09.98)</td><td>US</td></tr> <tr><td>Not furnished</td><td>22 September 1998 (22.09.98)</td><td>US</td></tr> <tr><td>60/069,391</td><td>4 November 1998 (04.11.98)</td><td>US</td></tr> <tr><td>60/128,660</td><td>8 April 1999 (08.04.99)</td><td>US</td></tr> </table> (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications <table> <tr><td>US</td><td>Not furnished (CIP)</td></tr> <tr><td>Filed on</td><td>17 September 1998 (17.09.98)</td></tr> <tr><td>US</td><td>09/158,720 (CIP)</td></tr> <tr><td>Filed on</td><td>22 September 1998 (22.09.98)</td></tr> <tr><td>US</td><td>Not furnished (CIP)</td></tr> <tr><td>Filed on</td><td>22 September 1998 (22.09.98)</td></tr> <tr><td>US</td><td>09/186,815 (CIP)</td></tr> <tr><td>Filed on</td><td>4 November 1998 (04.11.98)</td></tr> <tr><td>US</td><td>09/156,039 (CIP)</td></tr> <tr><td>Filed on</td><td>17 September 1998 (17.09.98)</td></tr> <tr><td>US</td><td>Not furnished (CIP)</td></tr> <tr><td>Filed on</td><td>4 November 1998 (04.11.98)</td></tr> <tr><td>US</td><td>60/128,660 (CIP)</td></tr> </table>		60/155,246	17 September 1998 (17.09.98)	US	09/158,720	22 September 1998 (22.09.98)	US	Not furnished	22 September 1998 (22.09.98)	US	60/069,391	4 November 1998 (04.11.98)	US	60/128,660	8 April 1999 (08.04.99)	US	US	Not furnished (CIP)	Filed on	17 September 1998 (17.09.98)	US	09/158,720 (CIP)	Filed on	22 September 1998 (22.09.98)	US	Not furnished (CIP)	Filed on	22 September 1998 (22.09.98)	US	09/186,815 (CIP)	Filed on	4 November 1998 (04.11.98)	US	09/156,039 (CIP)	Filed on	17 September 1998 (17.09.98)	US	Not furnished (CIP)	Filed on	4 November 1998 (04.11.98)	US	60/128,660 (CIP)	Filed on (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240, Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, Boulder Creek, CA 95006 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US). (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
60/155,246	17 September 1998 (17.09.98)	US																																									
09/158,720	22 September 1998 (22.09.98)	US																																									
Not furnished	22 September 1998 (22.09.98)	US																																									
60/069,391	4 November 1998 (04.11.98)	US																																									
60/128,660	8 April 1999 (08.04.99)	US																																									
US	Not furnished (CIP)																																										
Filed on	17 September 1998 (17.09.98)																																										
US	09/158,720 (CIP)																																										
Filed on	22 September 1998 (22.09.98)																																										
US	Not furnished (CIP)																																										
Filed on	22 September 1998 (22.09.98)																																										
US	09/186,815 (CIP)																																										
Filed on	4 November 1998 (04.11.98)																																										
US	09/156,039 (CIP)																																										
Filed on	17 September 1998 (17.09.98)																																										
US	Not furnished (CIP)																																										
Filed on	4 November 1998 (04.11.98)																																										
US	60/128,660 (CIP)																																										
		Published <i>Without international search report and to be republished upon receipt of that report.</i>																																									
(54) Title: RNA-ASSOCIATED PROTEINS																																											
(57) Abstract The invention provides human RNA-associated proteins (RNAAP) and polynucleotides which identify and encode RNAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RNAAP.																																											

(54) Title: RNA-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human RNA-associated proteins (RNAAP) and polynucleotides which identify and encode RNAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RNAAP.

1	M S R T Y L R P F M T S L V F V R N V A D D T R S E D L R R E K	399781
1	- G R Y G P I V D V V Y V F L D F T Y T R R F A G C F A Y V Q O F E D	GX 2961149
32	G R Y G P I V D V V Y V P L D Y T T R R F A G C F A Y V Q O F E D	399781
61	V R D A E D A L S M L D R K W I C G R O E I J G P A G G D P I	399781
61	V R D A E D A L S M L D R K W I C G R O E I J G P A G G D P I	GX 2961149
91	K T P H Q M K A K E G R A H V Y S S S R Y D D Y D R Y E R E S H	399781
91	K T P H Q M K A K E G R A H V Y S S S R Y D D Y D R Y E R E S H	GX 2961149
121	S R S Y S R R S R S R S R F D Y H Y R B Y S P R M S R P T	399781
121	S R S Y S R R S R S R S R F D Y H Y R B Y S P R M S R P T	GX 2961149
151	G R P R E R R E A I F Y T H I D Q T A A G I T S T V L L Y T L G	399781
151	G R P R E R R E A I F Y T H I D Q T A A G I T S T V L L Y T L G	GX 2961149
181	E R S T L S G S K R T K E G O F Y K R R F K G G N X V U L Q Y F -	399781
157	- E R S T L S G S K R T K E G O F Y K R R F K G G N X V U L Q Y F S A Y	GX 2961149
209	C T H I S I L T L V	399781
170	T P S - R K V	GX 2961149

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

RNA-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of RNA-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, immune/inflammatory, and reproductive disorders.

5

BACKGROUND OF THE INVENTION

Ribonucleic acid (RNA) is a linear single-stranded polymer of four ribonucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. 10 Ribosomal RNAs (rRNAs) are structural RNAs that are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that function in the translation of mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA 15 translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

RNA-binding proteins are essential for a wide variety of cellular and developmental
20 functions. They participate in RNA processing, editing, transport, localization, stabilization, and the posttranscriptional control of mRNAs. They also provide the protein component of ribosomal RNA, transfer RNA, and other ribonuclear proteins. The RNA binding activity of these proteins is mediated by specific RNA-binding domains contained within the proteins. A variety of conserved RNA binding motifs have been defined through comparisons of amino acid homologies and
25 structural similarities within these RNA-binding domains. These motifs include the RNP motif, an arginine-rich motif, the zinc-finger motif, the Y-box, the KH motif, and the double-stranded RNA-binding domain (dsRBD), all of which are characterized by specific consensus sequences (Burd, C. G. and Dreyfuss, G. (1994) Science 265:615 - 621).

RNA Processing

30 Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The spliceosomal complex is comprised of five small

nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) Biochemistry W.H. Freeman and Company, New York NY, p. 863).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have roles in functions that include splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) *Clin. Exp. Rheumatol.* 16:317-326). Some examples of hnRNPs include the yeast 10 proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) *Genes Dev.* 12:679-691). A common feature of all of these RNA-binding proteins is a glycine-rich region in the form of RGG repeats. HnRNPs have been shown to be important targets of the autoimmune response in 15 rheumatic diseases (Biamonti et al., supra).

An important means of regulating the function of hnRNPs is by methylation of arginine residues. The hnRNPs contain 65% of the methylated arginine residues in the cell nucleus. Methylation occurs within the RGG domain. Methylated arginine residues are also found in non-hnRNP RNA-binding proteins, all of which contain RGG repeats. The yeast enzyme, Hmt1p, is 20 responsible for methylation of Npl3p and Hrp1p. In HMT1 null mutants, methylation of these proteins is not detectable, and poly(A⁺)RNA accumulates in the nucleus. A molecular model predicts that Cbp80, Npl3p, and Hrp1p form a complex with mRNA to package the RNA for export from the nucleus, and that methylation plays a role in the efficiency of this packaging. Formation of this export complex is crucial for efficient exit of mRNA out of the nucleus. (Shen, 25 supra.) A human homolog of Hmt1p, HRMT1L2, has been identified and is required for methylation of arginine residues in the RGG repeats of hnRNP A1. (Scott, H.S. et al. (1998) *Genomics* 48:330-340.) Also, viral RNA-binding proteins, such as the herpes simplex virus ICP27 protein, are known to be arginine-methylated. This exploitation of the cellular export machinery may facilitate maturation of viral RNAs. (Shen, supra.)

30 Human myxoid liposarcomas have been shown to contain a chromosomal translocation [(t12;16)(q13;p11)] wherein the gene coding for an inhibitory, growth arrest-associated transcription factor, known as CHOP (C/EBP homologous protein), is fused to the gene for TLS (translocated in liposarcoma), a nuclear RNA-binding protein that contains an RNP motif. TLS has been shown to function as an RNA chaperone, shuttling RNA into and out of the nucleus

(Zinszner, H. et al. (1997) J. Cell Sci. 110:1741-1450). The fusion of TLS with CHOP serves to convert a transcription factor involved in growth arrest into one associated with abnormal cell proliferation (Crozat, A. et al. (1993) Nature 363:640-644). Subsequent work has shown that TLS and its homologs (e.g., EWS, associated with Ewing's sarcoma) comprise the N-terminal portion 5 of a number of fusion oncoproteins associated with sarcomas as well as with certain human acute myeloid leukemias (AMLs), secondary AMLs associated with myelodysplastic syndrome, and certain chronic myeloid leukemias (Aman, P. et al. (1996) Genomics 37:1-8; Zinszner, H. et al. (1997) Oncogene 14:451-461; Pereira, D.S. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8239-8244).

10 Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM) (Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816). The RRM is about 80 amino acids in length and forms four β -strands and two α -helices arranged in an α/β sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include 15 heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as Drosophila melanogaster and Caenorhabditis elegans. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively (Hodgkin, J. et al. (1994) 20 Development 120:3681-3689).

RNA Stability and Degradation

RNA helicases alter and regulate RNA conformation and secondary structure by using energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the "DEAD-box family," so named for 25 the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in various processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and 30 embryogenesis. All DEAD-box helicases contain several conserved sequence motifs within about 420 amino acids. These motifs include an A-type ATP binding motif, the DEAD-box/B-type ATP-binding motif, a serine/arginine/threonine tripeptide of unknown function, and a C-terminal glycine-rich motif with a possible role in substrate binding and unwinding. In addition, alignment of divergent DEAD-box helicase sequences has shown that 37 amino acid residues are identical

among these sequences, suggesting that conservation of these residues is important for helicase function. (Reviewed in Linder, P. et al. (1989) *Nature* 337:121-122.) Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors, suggesting that DDX1 may promote or enhance tumor progression by 5 altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in ultraviolet light-induced tumors, B-cell lymphoma, and myeloid malignancies (Godbout, R. et al. (1998) *J. Biol. Chem.* 273:21161-21168).

Ribonucleases (RNases) catalyze the hydrolysis of phosphodiester bonds in RNA chains, 10 thus cleaving the RNA. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid, which occurs in cells invaded by retroviruses. RNase H is an important enzyme in the retroviral replication cycle. RNase H domains are often found associated with reverse transcriptases. RNase activity in serum and cell extracts is elevated in a variety of cancers 15 and infectious diseases (Schein, C.H. (1997) *Nat. Biotechnol.* 15:529-536). Regulation of RNase activity may be a means for controlling tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

Translation

Proteins are translated from their RNA templates on the ribosome. The eukaryotic 20 ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, the ribosome also contains more than fifty proteins. The ribosomal proteins have a prefix which denotes the subunit to which they belong, either L (large) or S (small). Three important sites are identified on the ribosome: i) the aminoacyl-tRNA site (A site) where charged tRNAs (except the initiator-tRNA) bind on arrival; ii) 25 the peptidyl-tRNA site (P site) where new peptide bonds are formed and where the initiator tRNA binds, and iii) the exit site (E site) where deacylated tRNAs bind prior to their release from the ribosome (see Stryer, L. (1995) *Biochemistry* W.H. Freeman and Company, New York NY pp. 875-908; and Lodish, H. et al. (1995) *Molecular Cell Biology* Scientific American Books, New York NY pp. 119-138).

tRNA Charging

An important family of RNA-processing enzymes in the cytoplasm is the aminoacyl-transfer RNA (tRNA) synthetases. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes

can be divided into two structural classes, each class characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel β -sheet motif, as well as N- and C-terminal regulatory domains. Class II
5 enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains. (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530.)

One of the best studied of the aminoacyl-tRNA synthetases is seryl-tRNA synthetase (SerRS). SerRS is a class II enzyme with an N-terminal regulatory domain in the form of a
10 solvent exposed, antiparallel coiled-coil (the "helical arm"). A multiple sequence alignment and similarity plot of SerRS enzymes from prokaryotes, such as *E. coli*, and eukaryotes, such as yeast and mice, demonstrate the greatest variability in the N-terminal helical arm domain. Eukaryotic SerRS enzymes also contain a 20-48 amino acid C-terminal extension not found in prokaryotic synthetases. Truncation of the N-terminal helical arm causes SerRS to lose specificity for serine-
15 tRNA, such that the truncated SerRS reacts with non-cognate tRNAs as well. In eukaryotes, loss of the C-terminal sequence does not have a major affect on enzymatic activity. (Hartlein, *supra*; and Weygand-Duraševic, I. et al. (1996) J. Biol. Chem. 271:2455-2461.)

Autoantibodies against aminoacyl-tRNAs are generated by patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These
20 antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Translation Initiation

Initiation of translation can be divided into three stages. First an initiator transfer RNA (Met-tRNA_i) joins the 40S ribosomal subunit to form the 43S preinitiation complex. Next the 43S
25 preinitiation complex binds the mRNA, and migrates to the correct AUG initiation codon. In the third step, the 60S ribosomal subunit joins the 40S subunit to generate an 80S ribosome at the initiation codon. Regulation of translation primarily involves the first and second stage in the initiation process (V.M. Pain (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in
30 bringing an initiator tRNA and 40S ribosomal subunit together. eIF2B, a guanine nucleotide exchange protein, converts eIF2 from its GDP-bound inactive form to its GTP-bound active form. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA, bound to GTP, to the 40S ribosomal subunit. Two other factors, eIF1A and eIF3, bind and stabilize the 40S subunit by interacting with 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also

WO 00/15799

involved in association of the 40S ribosomal subunit with mRNA. The Met-tRNA_f, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, *supra*).

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m⁷GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide.

5 eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the

10 mRNA (M.W. Hentze (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by two structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that impede binding of the 43S preinitiation complex. Interestingly, the group of 15 mRNAs possessing highly structured 5' UTRs includes a disproportionately high number of mRNAs encoding proteins that take part in or regulate processes involved in cell proliferation. The efficiency with which these mRNAs are translated may play a crucial role in the maintenance of correct restraints on cell growth. Additionally, regulatory proteins may bind to sites within the 5' UTR and stabilize this secondary structure to prevent translation. The helicase activity of eIF4A 20 is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, *supra*).

The second structural feature of mRNA regulating binding of the 43S preinitiation complex is the 3' poly(A) tail. The translational efficiency of an mRNA is related to the length of its poly(A) tail, such that the longer the tail the more efficient the translation of the message. This 25 is due to an interaction between a protein that binds the poly(A) tail, the poly(A)-binding protein (PABP), and eIF4G. This interaction between PABP and eIF4G can only occur in the presence of RNA and involves a <120 amino acid site in the C-terminal half of eIF4G. This is an important form of regulation in translation of maternally-derived messages in early embryogenesis. The egg contains numerous mRNA molecules. Molecules with long poly(A) tails are translated early in 30 development and then undergo poly(A) tail shortening to repress further translation. Messages with short poly(A) tails, which are initially left untranslated, go through a cytoplasmic tail elongation to initiate translation later in development. This process of tail length modification responds to developmental cues and also appears to involve PABP (Pain, *supra*).

Another level of regulation involving eIF4G has been demonstrated by infection of

mammalian cells with picornaviruses. Several members of the picornavirus family, including poliovirus, human rhinovirus 2, and foot-and-mouth disease virus, inhibit cellular mRNA translation by cleaving eIF4G into two fragments. This cleavage by the viral protease effectively separates the N-terminal eIF4E binding site from the C-terminal binding sites for eIF4A, eIF3, and 5 PABP. Picornavirus RNAs, which are uncapped, utilize the C-terminal fragment of eIF4G for translation. This C-terminal fragment contains a region that interacts, either directly or indirectly, with an internal ribosome entry site (IRES) on the viral RNA molecule. Thus, eIF4G acts as a bridge between the 40S ribosome and the viral IRES for cap-independent translation as well (Hentze, *supra*).

10 Recently, a protein (p97) in yeast was shown to resemble the C-terminal fragment of eIF4G produced by picornavirus protease cleavage. p97 binds to both eIF3 and eIF4A, and may be involved in cap-independent translation of cellular mRNAs, though no candidate RNA species have been found within eukaryotic cells. p97 has been shown to be involved in modulating γ -interferon-induced programmed cell death (Hentze, *supra*).

15 Translation Elongation

Elongation, the joining of additional amino acids to the initiator methionine to complete the polypeptide chain, involves elongation factors EF1 α , EF1 β γ , and EF2. EF1 α is a GTP-binding protein which, when bound by GTP, brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the 20 initiation methionine. The GTP on EF1 α is hydrolyzed to GDP, and EF1 α -GDP dissociates from the ribosome. EF1 β - γ binds EF1 α -GDP and induces the dissociation of GDP from EF1 α , allowing EF1 α to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site 25 of the ribosome.

Translation Termination

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

The discovery of new RNA-associated proteins and the polynucleotides encoding them 30 satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, immune/inflammatory, and reproductive disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, RNA-associated proteins,

WO 00/15799

referred to collectively as "RNAAP" and individually as "RNAAP-1," "RNAAP-2," "RNAAP-3," "RNAAP-4," "RNAAP-5," "RNAAP-6," "RNAAP-7," "RNAAP-8," "RNAAP-9," "RNAAP-10," "RNAAP-11," "RNAAP-12," "RNAAP-13," "RNAAP-14," "RNAAP-15," "RNAAP-16," and "RNAAP-17." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and fragments thereof.

5 The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides an isolated and purified 10 polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

15 Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino 20 acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the 25 hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at 30 least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

5 The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

10 The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

15 The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

20 The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

25 The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

25 BRIEF DESCRIPTION OF FIGURES AND TABLES

Figure 1 shows the amino acid sequence alignment between RNAAP-1 (Incyte Clone number 399781; SEQ ID NO:1) and the human TLS-associated protein TASR (GI 2961149; SEQ ID NO:35), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

30 Figures 2A-H show the amino acid sequence alignment between RNAAP-2 (1252206; SEQ ID NO:2) and human eIF4G1 (GI 2660712; SEQ ID NO:36), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 3A and 3B show the hydropathy plots of RNAAP-2 (1252206; SEQ ID NO:2) and

human eIF4G1 (GI 2660712; SEQ ID NO:36), respectively. Plots were produced using MACDNASIS PRO software (Hitachi Software Engineering, S. San Francisco CA).

Figures 4A and 4B show the amino acid sequence alignment between RNAAP-3 (2950994; SEQ ID NO:3) and Drosophila seryl-tRNA synthetase (GI 2440051; SEQ ID NO:37), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 5A-C show the amino acid sequence alignment between RNAAP-4 (3461657; SEQ ID NO:4) and human arginine methyltransferase (GI 1808648; SEQ ID NO:38), produced using the multisequence alignment program of LASERGENE software.

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding RNAAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of RNAAP.

Table 3 shows useful fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze RNAAP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

10 DEFINITIONS

“RNAAP” refers to the amino acid sequences of substantially purified RNAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

15 The term “agonist” refers to a molecule which, when bound to RNAAP, increases or prolongs the duration of the effect of RNAAP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of RNAAP.

An “allelic variant” is an alternative form of the gene encoding RNAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered 20 mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

25 “Altered” nucleic acid sequences encoding RNAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as RNAAP or a polypeptide with at least one functional characteristic of RNAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding RNAAP, and improper or 30 unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding RNAAP. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent RNAAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity,

hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of RNAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values 5 may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of RNAAP which are preferably at least 5 to about 15 amino acids 10 in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of RNAAP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated 15 with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to RNAAP, decreases the 20 amount or the duration of the effect of the biological or immunological activity of RNAAP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of RNAAP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies 25 that bind RNAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, 30 and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on

the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules 5 may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or 10 biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic RNAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of 15 polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength 20 of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given 25 polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding RNAAP or fragments of RNAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), 30 detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping

sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of 5 the presence of nucleic acids, the same or related to a nucleic acid sequence encoding RNAAP, by northern analysis is indicative of the presence of nucleic acids encoding RNAAP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding RNAAP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

10 The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any 15 similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

20 The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of 25 hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). 30 In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR,

Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) Parameters for each method may be the default parameters provided by MEGALIGN or may be specified by the user. The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity.

Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term “humanized antibody” refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term “hybridization complex” refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” and “addition” refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which

may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

5 The term "modulate" refers to a change in the activity of RNAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of RNAAP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to 10 DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:18-34, for example, as distinct from any other sequence in the same genome. For 15 example, a fragment of SEQ ID NO:18-34 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:18-34 from related polynucleotide sequences. A fragment of SEQ ID NO:18-34 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:18-34 and the region of SEQ ID NO:18-34 to which the fragment corresponds are routinely determinable by one of ordinary skill 20 in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" and "operably linked" refer to functionally related 25 nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

30 The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding RNAAP, or fragments thereof, or RNAAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to

- various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment.
- 5 The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of RNAAP polypeptides refers to an amino acid sequence that is altered by 10 one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, 15 inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to RNAAP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice 20 variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting 25 polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

30 **THE INVENTION**

The invention is based on the discovery of new human RNA-associated proteins (RNAAP), the polynucleotides encoding RNAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immune/inflammatory, and reproductive disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding RNAAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each RNAAP were identified, and column 4 shows the 5 cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The clones in column 5 were used to assemble the consensus nucleotide sequence of each RNAAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the 10 invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows the identity of each polypeptide; and column 7 shows analytical methods used to identify each polypeptide through sequence homology and protein 15 motifs. The segment of RNAAP-1 from residue R51 through residue D60, corresponding to region BL00030B, received a score of 1118 on a strength of 1104, while the segment from residue L12 through residue F30, corresponding to region BL00030A, received a score of 1089 on a strength of 1095, and supported the presence of BL00030B with a *P* value less than 2.4×10^{-4} .

As shown in Figure 1, RNAAP-1 has chemical and structural similarity with the human 20 TLS-associated protein, TASR (GI 2961149; SEQ ID NO:35). In particular, RNAAP-1 and TASR share 76% identity, including the RNA recognition motif.

As shown in Figures 2 A-H, RNAAP-2 has chemical and structural similarity with human 25 eIF4G1 (GI 2660712; SEQ ID NO:36). In particular, RNAAP-2 and human eIF4G1 share 45% identity and have similar isoelectric points (5.23 and 5.04, respectively). As shown in Figures 3A and 3B, RNAAP-2 and human eIF4G1 have similar hydrophobicity profiles.

As shown in Figures 4A and 4B, RNAAP-3 has chemical and structural similarity with Drosophila seryl-tRNA synthetase (GI 2440051; SEQ ID NO:37). In particular, RNAAP-3 and 30 seryl-tRNA synthetase share 41% identity.

As shown in Figures 5A, 5B, and 5C, RNAAP-4 has chemical and structural similarity 35 with human arginine methyltransferase (GI 1808648; SEQ ID NO:38). In particular, RNAAP-4 and arginine methyltransferase share 46% identity.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding RNAAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1.

These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:18-34 and to distinguish between SEQ ID NO:18-34 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express RNAAP as a fraction of total tissues expressing RNAAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing RNAAP as a fraction of total tissues expressing RNAAP. Northern analysis shows the expression of SEQ ID NO:18 in various libraries, at least 51% of which are associated with cancer and at least 29% of which are associated with inflammation and the immune response. Of particular note is SEQ ID NO: 29, which is expressed in only 25 libraries, 10 10(40%) of which are associated with reproductive tissue and 17(76%) of which are associated with cell proliferative disorders. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses RNAAP variants. A preferred RNAAP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the RNAAP amino acid sequence, and which contains 20 at least one functional or structural characteristic of RNAAP.

The invention also encompasses polynucleotides which encode RNAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34, which encodes RNAAP.

The invention also encompasses a variant of a polynucleotide sequence encoding RNAAP. 25 In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding RNAAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34 which has at least about 70%, more preferably at least about 85%, and most 30 preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:18-34. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of RNAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the

genetic code, a multitude of polynucleotide sequences encoding RNAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These 5 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring RNAAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode RNAAP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring RNAAP under 10 appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding RNAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for 15 substantially altering the nucleotide sequence encoding RNAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode RNAAP 20 and RNAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding RNAAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of 25 hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:18-34 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium 30 citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and

most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, 5 hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions 10 will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM 15 NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 20 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow 25 fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system 30 (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art.

(See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding RNAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal

using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

5 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode RNAAP may be cloned in recombinant DNA molecules that direct expression of RNAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express
10 RNAAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter RNAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments
15 and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding RNAAP may be synthesized, in whole or in
20 part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:225-232.) Alternatively, RNAAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis
25 may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of RNAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid
30 chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active RNAAP, the nucleotide sequences encoding

RNAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding RNAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding RNAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding RNAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding RNAAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding RNAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding RNAAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding RNAAP can

be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding RNAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these 5 vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of RNAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of RNAAP may be used. For example, vectors containing the strong, inducible T5 or T7 10 bacteriophage promoter may be used.

Yeast expression systems may be used for production of RNAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable 15 integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) *Methods Enzymol.* 153:516-54; and Scorer, C. A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of RNAAP. Transcription of sequences encoding RNAAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used 20 alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA 25 transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding RNAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite 30 leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses RNAAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat.*

5 *Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of RNAAP in cell lines is preferred. For example, sequences encoding RNAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate 10 vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

15 Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* or *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers 20 resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., 25 Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol.*

30 *Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding RNAAP is inserted within a marker gene sequence, transformed cells containing sequences encoding RNAAP can be identified by the absence of

marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding RNAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding RNAAP and that 5 express RNAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

10 Immunological methods for detecting and measuring the expression of RNAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on RNAAP is preferred, 15 but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

20 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding RNAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding RNAAP, or any fragments thereof, may be 25 cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter 30 molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding RNAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The

protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode RNAAP may be designed to contain signal sequences which direct secretion of RNAAP through a prokaryotic or eukaryotic cell membrane.

- 5 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity.
- 10 Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding RNAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric RNAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of RNAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the RNAAP encoding sequence and the heterologous protein sequence, so that RNAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled RNAAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems

(Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S -methionine.

- Fragments of RNAAP may be produced not only by recombinant production, but also by
- 5 direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, *supra*, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of RNAAP may be synthesized separately and then combined to produce the full length molecule.

10 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of RNAAP and RNA-associated proteins. In addition, the expression of RNAAP is closely associated with reproductive tissues, nervous tissues, cell proliferation including cancer, and inflammation and immune response. Therefore, RNAAP appears to play a role in cell

15 proliferative, immune/inflammatory, and reproductive disorders. In the treatment of disorders associated with increased RNAAP expression or activity, it is desirable to decrease the expression or activity of RNAAP. In the treatment of the above conditions associated with decreased RNAAP expression or activity, it is desirable to increase the expression or activity of RNAAP.

Therefore, in one embodiment, RNAAP or a fragment or derivative thereof may be

20 administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera,

25 psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an

30 immune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes

mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis,

5 pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin

10 production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis,

15 cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing RNAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those described above.

20 In a further embodiment, a pharmaceutical composition comprising a substantially purified RNAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of RNAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those listed above.

25 In a further embodiment, an antagonist of RNAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, those described above. In one aspect, an antibody

30 which specifically binds RNAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express RNAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding RNAAP may be administered to a subject to treat or prevent a disorder associated with

increased expression or activity of RNAAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination 5 therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

10 An antagonist of RNAAP may be produced using methods which are generally known in the art. In particular, purified RNAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind RNAAP. Antibodies to RNAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab 15 fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with RNAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various 20 adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

25 It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to RNAAP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of 30 RNAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to RNAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-

hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the
5 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate
antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984)
Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and
Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the
production of single chain antibodies may be adapted, using methods known in the art, to produce
10 RNAAP-specific single chain antibodies. Antibodies with related specificity, but of distinct
idiotypic composition, may be generated by chain shuffling from random combinatorial
immunoglobulin libraries. (See, e.g., Burton D.R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte
population or by screening immunoglobulin libraries or panels of highly specific binding reagents
15 as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86:
3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for RNAAP may also be
generated. For example, such fragments include, but are not limited to, F(ab')2 fragments
produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing
20 the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be
constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired
specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the
desired specificity. Numerous protocols for competitive binding or immunoradiometric assays
25 using either polyclonal or monoclonal antibodies with established specificities are well known in
the art. Such immunoassays typically involve the measurement of complex formation between
RNAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing
monoclonal antibodies reactive to two non-interfering RNAAP epitopes is preferred, but a
competitive binding assay may also be employed (*Pound, supra*).

30 Various methods such as Scatchard analysis in conjunction with radioimmunoassay
techniques may be used to assess the affinity of antibodies for RNAAP. Affinity is expressed as
an association constant, K_a , which is defined as the molar concentration of RNAAP-antibody
complex divided by the molar concentrations of free antigen and free antibody under equilibrium
conditions. The K_a determined for a preparation of polyclonal antibodies, which are

heterogeneous in their affinities for multiple RNAAP epitopes, represents the average affinity, or avidity, of the antibodies for RNAAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular RNAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the RNAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of RNAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of RNAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding RNAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding RNAAP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding RNAAP. Thus, complementary molecules or fragments may be used to modulate RNAAP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding RNAAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding RNAAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding RNAAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding RNAAP.

Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

- As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding RNAAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred.
- 10 Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al.
- 15 (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For 20 example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding RNAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 25 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be 30 prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding RNAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or

SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by 10 endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers 15 may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) *Nature Biotechnology* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

20 An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of RNAAP, antibodies to RNAAP, and mimetics, agonists, antagonists, or inhibitors of RNAAP. The compositions may be administered alone or in combination with at least one other agent, such 25 as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any 30 number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which

facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

- Pharmaceutical compositions for oral administration can be formulated using 5 pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

- Pharmaceutical preparations for oral use can be obtained through combining active 10 compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, 15 including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

- Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, 20 polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

- Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. 25 Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

- Pharmaceutical formulations suitable for parenteral administration may be formulated in 30 aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include

fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

5 For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

10 The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2%
15 sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of RNAAP, such labeling would include amount, frequency, and method of administration.

20 Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in
25 cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example
30 RNAAP or fragments thereof, antibodies of RNAAP, and agonists, antagonists or inhibitors of RNAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic

effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind RNAAP may be used for the diagnosis of disorders characterized by expression of RNAAP, or in assays to monitor patients being treated with RNAAP or agonists, antagonists, or inhibitors of RNAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for RNAAP include methods which utilize the antibody and a label to detect RNAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring RNAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of RNAAP expression. Normal or standard values for RNAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to

RNAAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of RNAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the 5 parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding RNAAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of RNAAP 10 may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of RNAAP, and to monitor regulation of RNAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting 15 polynucleotide sequences, including genomic sequences, encoding RNAAP or closely related molecules may be used to identify nucleic acid sequences which encode RNAAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding RNAAP, allelic variants, or related sequences.

20 Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the RNAAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:18-34 or from genomic sequences including promoters, enhancers, and introns of the RNAAP gene.

25 Means for producing specific hybridization probes for DNAs encoding RNAAP include the cloning of polynucleotide sequences encoding RNAAP or RNAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled 30 by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding RNAAP may be used for the diagnosis of disorders associated with expression of RNAAP. Examples of such disorders include, but are not limited to,

a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding RNAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered RNAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding RNAAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The

nucleotide sequences encoding RNAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly 5 altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding RNAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of 10 RNAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding RNAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially 15 purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, 20 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 25 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding 30 RNAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding RNAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding RNAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less

stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

- Methods which may also be used to quantify the expression of RNAAP include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding RNAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding RNAAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene

sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, RNAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between RNAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with RNAAP, or fragments thereof, and washed. Bound RNAAP is then detected by methods well known in the art. Purified RNAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding RNAAP specifically compete with a test compound for binding RNAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with RNAAP.

In additional embodiments, the nucleotide sequences which encode RNAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

5 The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0598 P, filed September 22, 1998], U.S. Ser. No. [Attorney Docket No. PF-0600 P, filed September 17, 1998], U.S. Ser. No. [Attorney Docket No. PF-0626 P, filed November 4, 1998], and U.S. Ser. No. 60/128,660, are hereby expressly incorporated by reference.

10 **EXAMPLES**

I. **Construction of cDNA Libraries**

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

20 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

25 In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the

polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies. II.

5 **Isolation of cDNA Clones**

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and 15 thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. **Sequencing and Analysis**

20 cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in 25 ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other 30 sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those

skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein 5 in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were 10 generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, 15 dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open 20 reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. 25 HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Str. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:18-34. Fragments from about 20 to about 4000 nucleotides which are useful in 30 hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7;

Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the 5 sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the 10 length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in 15 which the transcript encoding RNAAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number 20 of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of RNAAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:18-34 were produced by extension 25 of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal 30 to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art.

PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer

sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:18-34 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, 5 and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:18-34 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide 10 fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). 15 An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 20 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array 25 elements on the surface of a substrate. (See, e.g., Baldeschweiler, *supra*.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels 30 and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected

using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g.,

5 UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

10 Sequences complementary to the RNAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring RNAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of

15 RNAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the RNAAP-encoding transcript.

IX. Expression of RNAAP

20 Expression and purification of RNAAP is achieved using bacterial or virus-based expression systems. For expression of RNAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (lac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express RNAAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of RNAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding RNAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional

genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, RNAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, 5 single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from RNAAP at specifically engineered sites. FLAG, an 8-amino acid 10 peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified RNAAP obtained by these methods can be used directly in the following activity assay.

15 X. Demonstration of RNAAP Activity

RNAAP activity is demonstrated by a polyacrylamide gel mobility-shift assay. In preparation for this assay, RNAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing RNAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to 20 allow expression and accumulation of RNAAP. Extracts containing solubilized proteins can be prepared from cells expressing RNAAP by methods well known in the art. Portions of the extract containing RNAAP are added to [³²P]-labeled RNA. Radioactive RNA can be synthesized in vitro by techniques well known in the art. The mixtures are incubated at 25 °C in the presence of RNase inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed 25 by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between RNAAP and the radioactive transcript. A band of similar mobility will be absent in samples prepared using control extracts prepared from untransformed cells.

Alternatively, the activity of RNAAP is measured as the level of *in vitro* translation of 30 cap-dependent chloramphenicol acetyltransferase (CAT) and cap-independent luciferase (LUC) reporter constructs (Haghishat, A., et al. (1996) J. Virol. 70:8444-8450). Bicistronic pGEMCAT/EMC/LUC mRNA is used in the assay. The first cistron on this mRNA construct encodes the CAT protein and its translation is cap-dependent. The second cistron encodes luciferase enzyme. The encoded region of the second cistron is preceded by the IRES of

encephalomyocarditis (EMC) virus, making luciferase translation cap independent. Linearized pGEMCAT/EMC/LUC is transcribed in vitro using T7 RNA polymerase in the presence of 10-fold molar excess m⁷GpppG, a cap analog that promotes capping of the RNA product. Rabbit reticulocyte lysate is treated with picornavirus 2A protease. Treatment of the lysate with 2A protease reduces cap-dependent (CAT) translation, but does not inhibit cap-independent (luciferase) translation. Treated lysate is programmed by addition of the capped mRNA in the presence of 20 µCi [³⁵S]methionine. Translation reaction mixtures are incubated for 90 min in the presence of added eIF4E, RNAAP, eIF4E and RNAAP, or with no additions. Translation products are analyzed by SDS-PAGE, acid fixation, and autoradiography. RNAAP activity is calculated based on the expression level of CAT relative to luciferase as compared to control reactions lacking RNAAP.

Alternatively, RNAAP activity is measured as the aminoacylation of a substrate tRNA in the presence of [¹⁴C]serine. RNAAP is incubated with tRNA^{ser} and [¹⁴C]serine in a buffered solution. ¹⁴C-labeled product is separated from free [¹⁴C]serine by chromatography, and the incorporated ¹⁴C is quantified by scintillation counter. The amount of ¹⁴C detected is proportional to the activity of RNAAP in this assay.

Alternatively, RNAAP activity is measured as the methylation of a substrate in the presence of [methyl-³H]-S-adenosylmethionine (SAM). RNAAP is incubated with an appropriate substrate and [methyl-³H]SAM in a buffered solution. ³H-labeled product is separated from free [methyl-³H]SAM by gel electrophoresis, and the incorporated ³H is quantified by fluorography. The amount of ³H detected is proportional to the activity of RNAAP in this assay.

XI. Functional Assays

RNAAP function is assessed by expressing the sequences encoding RNAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion

protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear 5 DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of 10 fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) *Flow Cytometry*, Oxford, New York NY.

The influence of RNAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding RNAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved 15 regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding RNAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

20 XII. Production of RNAAP Specific Antibodies

RNAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the RNAAP amino acid sequence is analyzed using LASERGENE 25 software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide 30 synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit

antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring RNAAP Using Specific Antibodies

Naturally occurring or recombinant RNAAP is substantially purified by immunoaffinity chromatography using antibodies specific for RNAAP. An immunoaffinity column is constructed by covalently coupling anti-RNAAP antibody to an activated chromatographic resin, such as 5 CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing RNAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of RNAAP (e.g., high ionic 10 strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/RNAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and RNAAP is collected.

XIV. Identification of Molecules Which Interact with RNAAP

RNAAP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter 15 reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RNAAP, washed, and any wells with labeled RNAAP complex are assayed. Data obtained using different concentrations of RNAAP are used to calculate values for the number, affinity, and association of RNAAP with the candidate molecules.

20 Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying 25 out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1 18	399781	PITUNOT02	399781H1 and 790764R1	399781X12 (PITUNOT02), 1271965F6 (TESTTUT02), 792124R1 (PROSTUT03), and 405935R1 (EOSIHET02)
2 19	1252206	LUNGFET03	1232931T6 (LUNGFET03), (BRSTNOT17), 3330287H1 (HEAONOT04), 1662596H1 (BRSTNOT09), 2266829R6 (UTRSNOT02), (BRAINOT14), (BRAWNOT01), SAJA00661R1,	3109423H1 (BRSTTUT15), 3269650H1 (BRAINOT20), 2655078H1 (THYMNOT04), 4333545H1 (KIDCTMT01), 078192R1 and 1252206F6 (LUNGFE03), 078192F1 (SYNORAB01), 1252206F6 (LUNGFE03), 1638473F6 (UTRSNOT06), SAJA00355F1, SAJA01106R1, SAJA01874F1, and SAJA02468F1
3 20	2950994	KIDNFET01	1968448H1 (LUNGNOT04), 3461657 293TF201	2606248F6 (LUNGUT07), 2052041X301D1 (LIVRFET02), 053076H1 (FIBRNOT01), 1968448H1 (BRSTNOT04), 2795721F6 (NPOLNOT01), (BRAUNOT02), 2789769F6 (COLNTUT16), SBUA03574D1 and SBUA00296D1
4 21				2052041X301D1 (LIVRFET02), 4341820F6 (293TF2T01), 3461657H1 (293TF2T01), 534171F1 (BRAINOT03), 4717220H1 (BRAIHTC02)
5 22	053076	FIBRNOT01		
6 23	1292379	PGANNOT03	458715T6 (KERANOT01), and 1292379T1 (PGANNOT03), 3447383H2 (BLADNOT09),	850050T1 (NGANNOT01), 1292379F1, 1292379H1 (PGANNOT03), 1398840F6 and 1398840T6 (BRAITUT08), 3780263H1 (BRSTNOT27)
7 24	1437783	PANCNOT08		117781F1 (KIDNNOT01), 1352071F1 (LATRTUT02), (PANCNOT08), 2527706H1 (GBLANOT02), 117781F1 (KIDNNOT01), 1352071F1 (LATRTUT02), 1437783H1 (PANCNOT08), 2527706H1 (GBLANOT02), 4567705H1 (HELATXT01)
8 25	1557635	BLADTUT04		077627R1 (SYNORAB01), 1568446F1 (UTRSNOT05), (TESTNOT03), 2098109H1 (BRAITUT02), 3866538H1 (BRAITUT07)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
9 26	2049352	LIVRFET02	078075R1 (SYNORAB01), 994247R6 (COLNNOT11), 1334674F6 (COLNNOT13), 2049352F6 and 2049352H1 (LIVRFET02), 3219182H1 (COLNNON03)	
10 27	2231663	PROSNOT16	307827H1 (HEARNOT01), 1455948F1 and 1455948R1 (COLNFET02), 2231663H1 (PROSNOT16), 3779128H1 (ERSTNOT27)	
11 28	2604449	LUNGUT07	606296R6 (BRSTTUT01), 1718568T6 (BLADNOT06), 2604449F6 and 2604449H1 (LUNGUT07), 5093027F6 (UTRSTMRO1), SAEA01050F1, SAEA01365F1, SAECA01108F1, SBKA0068LF1	
12 29	2604993	LUNGUT07	1441072F6 and 1441072T6 (THYRNOT03), 2604993H1 (LUNGUT07), 3389190T6 (LUNGUT17), SBIA05937D1, SBIA11687D1, SBIA04881D1, SBIA03937D1, SBIA00985D1	
13 30	2879070	UTRSTUT05	1458387F7, 1458387R1, and 1458387T6 (COLNFET02), 1858014X14C1 (PROSNOT18), 2595610H1 (OVARTUT02), 2879070H1 (UTRSTUT05)	
14 31	3093845	BRSTNOT19	134421R1 (BMARNOT02), 979683R6 (TONGTUT01), 3093845F6 and 3093845H1 (BRSTNOT19), 3294785F6 (TLYJINT01)	
15 32	3685685	HEAANOT01	1556450F1 (BLADTUT04), 1615712T6 (BRAITUT12), 2041291R6 (HIPONON02), 2448460F6 (THP1NOT03), 3685685H1 (HEAANOT01), 3954790H1 (PONSAZT01), 4918977H2 (TESTNOT11)	
16 33	3825977	BRAINOT23	2373839T6 and 2375912X302D1 (ISLTNOT01), 3825977H1 (BRAINOT23), 3882790H1 (SPLNNOT11), SBIA02579D1, SBIA02994D1, SBIA10082D1, SBIA06183D1, SBIA05526D1, SBIA02807D1	
17 34	4941262	BRAIFEN03	4941262F6 and 4941262H1 (BRAIFEN03)	

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites on Sites	Signature Sequence	Identification	Analytical Methods
1 216		S129, T21, S108, T161, T178, T47, S107, S143, T150, S185, Y116, Y138	N9	RNA recognition motif: L12-I83 RNA-binding region RNP-1 R51 signature: R1-D60, L12-F30	GI 2961149 Human TLS-associated protein, TASR	Motifs BLAST PFAM BLOCKS
2 1584		S740, S888, T294, S304, S370, T517, S584, S598, T615, S865, T1058, T1115, S1155, T1190, S1209, S1227, S1264, S1333, S1381, S1421, S1501, T1503, S1550, S30, T141, S362, S456, S611, S700, T817, S965, S985, S1121, T1126, T1144, S1155, T1175, S1200, S1333, S1367, S1381, S1416, T1286, S1381, S1550	N1162, S257, S317, S366, S542, S582, N1188, N1195, S718, S1085, S1164, S1217, S1290, S1416, T1503, S304, T491, T507, S718, S735, S985, S1121, T1144, S1155, T1286, S1381, S1550	Leucine zipper pattern: L1513-L1534 Wilms' tumor protein: G80-P94, S412-H426	GI 2660712 Human eIF4G1	Motifs BLAST PRINTS

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
3	166	S78, T135	N72, N99		GI 2440051 seryl-tRNA synthetase	BLAST
4	531	S27, T58, S59, S157, S242, S339, S428, S430, S242, T439, S475, S492, Y89	N155, N522, N523	C2H2 type zinc finger motif: C50-H71 N-methyltransferase cofactor-binding motif: V259-A273	GI 1808648 Human arginine methyltransferase	Motifs BLAST BLOCKS PFAM PRINTS
5	148	S32 S38 S47 T69 T141 Y60		A31-D115 (Ribosomal L27 protein) M1-A27 (Signal peptide)	ribosomal protein L27 9 642605	Motifs BLAST PFam HMM SPScan
6	317	S20 S40 S106 S110 S117 T135 T142 S144 T260 S302 S6 S10 T134 S215 S281	N148 N208 N228		pre-ribosomal particle assembly protein 9 2398808	Motifs BLAST
7	278	T10 S83 S56 T57 T61 T121 S202 S244 T13 T68 T156 T192 S224 Y251	N71 N120		translation initiation factor 3 (infC) 9 3844793	Motifs BLAST

Table 2 (cont.)

Seq ID No:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
8	586	T29 T81 T261 S512 T4 S21 T29 S97 S227 T229 S235 T348 S371 S417 T475 T485 S511 S513 S515 S554 T562 S77 T127 T194 S206 S215 S256 S356 S479 Y274 Y297 Y309	N427		Similar to mRNA splicing factor g 3878326	Motifs BLAST
9	384	T32 S167 T327 T339 T349 S28 T148 T311 S372 Y13 Y19 Y86 Y277	N229	H257-M296 (Cytidine and deoxycytidylate deaminases Zinc-binding region signature)	phorbolin I protein kinase C associated protein g 436941	Motifs BLAST
10	325	T61 S298 S320 S49 T53 S116	N163	R94-G302 (L1P family ribosomal proteins)	Ribonucleotide reductase subunit M2 g 200768	Motifs BLAST Pfam

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential glycosylation sites	Signature Sequence	Identification	Analytical Methods
11	351	S39 T182 S329 S18 S29 T65 T182 S225 S38 Y87	N23 N314	E131-I146 (Ribonucleotide reductase small subunit) P46-D100, F123-D148, F198-F239, V251-R292 (Ribonucleotide reductase) W69-Y331 (Ribonucleotide reductase) R186-W207 (transmembrane)	Ribonucleotide reductase subunit M2 g 200468	Motifs BLAST Pfam BLOCKS HMM
12	681	T68 S79 S135 T160 S179 S201 S216 S237 T301 T312 T338 T363 T405 T457 S524 S123	N89 N600 N623	V227-V297, V328-L401, I447-V520 (RNA recognition motif) M1-K22 (signal peptide)	Similarity to Human heterogeneous nuclear ribonucleoprotein (hnRNP) F protein g 3880146	Motifs BLAST Pfam SPScan
13	408	S3 S45 S68 T212 T236 S248 T145 T279 Y193	N206	I121-M144 (transmembrane)	RNA helicase A g2880057	Motifs BLAST HMM

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential glycosylation sites	Signature Sequence	Identification	Analytical Methods
14	351	S126 S5 T7 S75 S108 S140 S195 S314 S339 S59 S122 S254 S300 S344 Y23	N113 N202	K36-Y43 (Eukaryotic putative RNA-binding region RNP-1 signature) I2-L38, V127-V194, I269-V334 (RNA recognition motif)	He1-N2 RNA binding protein q905387	Motifs BLAST Pfam
15	472	S69 S116 S346 S89 S237 S239 S301 T303 S358 S4 T39 S124 T176	N219 N248	102-130, 178-204 (glycosyl hyrolase)	Human RNA binding protein q 2804465	Motifs BLAST PRINTS
16	616	S154 S368 S376 T570 S14 S44 T53 S83 S94 S466		V18-V89 (RNA recognition motif) E36-R85 (eukaryotic RNA-binding RNP-1)	Cleavage Stimulating factor q 181139	Motifs BLAST Pfam ProfileScan
17	112	T42 Y69		G74-P95 (ribosomal protein L35Ae signature) L12-F106 (ribosomal protein L35Ae signature)	q4392 ribosomal protein L37a	Motifs BLAST Pfam BLOCKS

Table 3

Polynucleotide SEQ ID NO:	Selected Fragment (Nucleotide number)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
18	30-90	Nervous (0.191) Reproductive (0.309)	Cell proliferation (0.510) Inflammation and Immune Response (0.290)	PSPORT1
19	1137-1196	Nervous (0.245) Reproductive (0.216)	Cell proliferation (0.560) Inflammation and Immune Response (0.230)	PINCY
20	454-510	Reproductive (0.263) Nervous (0.211)	Cancer (0.580) Inflammation and Immune Response (0.160)	PINCY
21	31-81	Nervous (0.357) Gastrointestinal (0.179) Reproductive (0.143)	Cancer (0.610) Inflammation and Immune Response (0.210)	PINCY
22	1-46	Reproductive (0.247) Nervous (0.183) Gastrointestinal (0.118)	Cell proliferation (0.613) Inflammation (0.290)	PBLUESCRIPT
23	273-317	Reproductive (0.256) Nervous (0.209)	Cell proliferation (0.465) Inflammation (0.256)	PINCY
24	434-478	Gastrointestinal (0.244) Nervous (0.186) Reproductive (0.163)	Cell proliferation (0.535) Inflammation (0.361)	PINCY
25	174-218	Reproductive (0.230) Nervous (0.216) Cardiovascular (0.122)	Cell proliferation (0.554) Inflammation (0.311)	PINCY

26	489-533	Reproductive (0.270) Hematopoietic/Immune (0.243) Nervous (0.162)	Cell proliferation (0.676) Inflammation (0.405)	pINCY
----	---------	---	--	-------

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragment (Nucleotide number)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
27	199-252	Reproductive (0.308) Cardiovascular (0.205)	Cell proliferation (0.770) Inflammation (0.128)	pINCY
28	110-154	Cardiovascular (0.289) Nervous (0.184) Reproductive (0.158)	Cell proliferation (0.685) Inflammation (0.158)	pINCY
29	326-370	Reproductive (0.400) Gastrointestinal (0.240) Cardiovascular (0.120)	Cell proliferation (0.760) Inflammation (0.240)	pINCY
30	516-563	Reproductive (0.415) Nervous (0.151) Hematopoietic/Immune (0.113)	Cell proliferation (0.566) Inflammation (0.320)	pINCY
31	272-316	Hematopoietic/Immune (0.286) Gastrointestinal (0.214) Reproductive (0.214)	Inflammation (0.714) Cell proliferation (0.495)	pINCY
32	119-163	Reproductive (0.328) Hematopoietic/Immune (0.219) Nervous (0.156)	Cell proliferation (0.672) Inflammation (0.313)	pINCY
33	812-856	Gastrointestinal (0.208) Hematopoietic/Immune (0.208) Developmental (0.167) Nervous (0.167)	Inflammation (0.541) Cell proliferation (0.458)	pINCY
34	42-86	Nervous (1.000)	Cell proliferation (1.000)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
18	PITUNOT02	Library was constructed using RNA isolated from the pituitary glands removed from a pool of 87 male and female donors, 15 to 75 years old (RNA acquired from Clontech, CLON 6584-1).
19	LUNGFET03	Library was constructed RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included bronchitis.
20	KIDNFET01	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian female fetus, who died at 17 weeks' gestation from anencephalus.
21	293TF201	Library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine (5AZA) and transformed with adenovirus 5 DNA.
22	FIBRN0T01	Library was constructed using RNA isolated from the WI38 lung fibroblast cell line, which was derived from a 3-month-old Caucasian female fetus. 2x10 ⁻⁶ primary clones were then amplified to stabilize the library for long-term storage.
23	PGANNOT03	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule.
24	PANCNOT08	Library was constructed using RNA isolated from pancreatic tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology for the associated tumor tissue indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Family history included cardiovacular disease, type II diabetes, and stomach cancer.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
25	BLADTUT04	Library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use. Family history included type I diabetes, malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.
26	LIVRFET02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included bronchitis.
27	PROSNOT16	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA) and was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
28	LUNGTUT07	Library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasiash male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.
29	LUNGTUT07	Library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasian male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
30	UTRSRSTU05	Library was constructed using RNA isolated from uterine tumor tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated uterine leiomyoma. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Patient history included a ventral hernia and a benign ovarian neoplasm.
31	BRSTNOT19	Library was constructed using RNA isolated from breast tissue removed from a 67-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated residual invasive lobular carcinoma. The focus of residual invasive carcinoma was positive for both estrogen and progesterone. Patient history included depressive disorder and benign large bowel neoplasm. Family history included cerebrovascular disease, benign hypertension, congestive heart failure, and lung cancer.
32	HEAANOT01	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco use. Family history included atherosclerotic coronary artery disease.

Table 4 (cont.)

Polymerase SEQ ID NO:	Library	Library Comment
33	BRAINOT23	Library was constructed using RNA isolated from right temporal lobe tissue removed from a 45-year-old Black male during a brain lobectomy. Pathology for the associated tumor tissue indicated dysembryoplastic neuroepithelial tumor of the right temporal lobe. The right temporal region dura was consistent with calcifying pseudotumor of the neuraxis. The patient presented with convulsive intractable epilepsy, partial epilepsy, and memory disturbance. Patient history included obesity, meningitis, backache, unspecified sleep apnea, acute stress reaction, acquired knee deformity, and chronic sinusitis. Family history included obesity, benign hypertension, cirrhosis of the liver, alcohol abuse, hyperlipidemia, cerebrovascular disease, and type II diabetes.
34	BRAIFEN03	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus with a hypoplastic left heart stillborn after 23 weeks' gestation. The library was normalized in two rounds (with 48 hour reannealing hybridizations) using conditions adapted from Soares et al. and Bonaldo et al.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	Mismatch <50%
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	<i>ESTs:</i> Probability value = 1.0E-8 <i>or less</i> <i>Full Length sequences:</i> Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfasix, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs:</i> fasta E value= 1.06E-6 <i>Assemblies/ ESTs:</i> fasta identity ~ 95% or greater and Match length=200 bases or greater; <i>fastx</i> E value=1.0E-8 or less <i>Full Length sequences:</i> fastx score=100 or greater
BLIMPS	A BLocks IMproved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff. Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribkoff, M. et al. (1988) CABIOS 4:61-66; Gribkoff, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score = 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:193-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score = 5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. ^{supra} ; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and fragments thereof.
2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.
7. A method for detecting a polynucleotide, the method comprising the steps of:
 - (a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and
 - (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.
8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.
9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ

ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90%
5 polynucleotide sequence identity to the polynucleotide of claim 9.
11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.
- 10 12. An expression vector comprising at least a fragment of the polynucleotide of
claim 3.
13. A host cell comprising the expression vector of claim 12.
14. A method for producing a polypeptide, the method comprising the steps of:
15. a) culturing the host cell of claim 13 under conditions suitable for the
expression of the polypeptide; and
b) recovering the polypeptide from the host cell culture.
15. A pharmaceutical composition comprising the polypeptide of claim 1 in
20 conjunction with a suitable pharmaceutical carrier.
16. A purified antibody which specifically binds to the polypeptide of claim 1.
- 25 17. A purified agonist of the polypeptide of claim 1.
18. A purified antagonist of the polypeptide of claim 1.
19. A method for treating or preventing a disorder associated with decreased
expression or activity of RNAAP, the method comprising administering to a subject in need of
30 such treatment an effective amount of the pharmaceutical composition of claim 15.
20. A method for treating or preventing a disorder associated with increased
expression or activity of RNAAP, the method comprising administering to a subject in need of
such treatment an effective amount of the antagonist of claim 18.

1	M S R Y L R P P N T S L F V R N V A D D T R S E D L R R E F	399781
1	M S R Y L R P P N T S L F V R N V A D D T R S E D L R R E F	GI 2961149
31	G R Y G P I V D V Y V P L D F Y T R R P R G F A Y V Q F E D	399781
31	G R Y G P I V D V Y V P L D F Y T R R P R G F A Y V Q F E D	GI 2961149
61	V R D A E D A L H N L D R K W I C G R Q I E I Q F A Q G D R	399781
61	V R D A E D A L H N L D R K W I C G R Q I E I Q F A Q G D R	GI 2961149
91	K T P N Q M K A K E G R N V Y S S S R Y D D Y D R Y R R S R	399781
91	K T P N Q M K A K E G R N V Y S S S R Y D D Y D R Y R R S R	GI 2961149
121	S R S Y E R R R S R S R S F D Y N Y R S Y S P R N S R P T	399781
121	S R S Y E R R R S R S R S F D Y N Y R S Y S P R N S R P T	GI 2961149
151	G R P R R R E A I P T M I D Q T A A G I P S T V L L T T L Q	399781
151	G R P R R R S -	GI 2961149
181	E R S E S G K R T K E G Q E K R P K G G W K V L Q Y E - - []	399781
157	- R S H S D N - - - - - - D R P N C S W N T - Q Y S S A Y	GI 2961149
209	C T N I L T L V	399781
178	Y T S - - R K I	GI 2961149

FIGURE 1

2/15

FIGURE 2A

3/15

211	A A S D Q K Q E E K P K P D P V L K S P S P V L R L V L S G	1252206
119	- - - - - - - - - P E P T P L - - - A E P T L E V E V T L	GI 2660712
241	E K K E Q E G O T S E T T A I V S I A E L P L P P S P T T V	1252206
136	S K P V P E S E F S S S - - - - - - - - - P L Q A P - - -	GI 2660712
271	S S V A R S T I A A P T S S A L S S O P I F T T A I D D R C	1252206
153	T P L A S H T V E T I H E P N G M V P S E D L E P E V E S S P	GI 2660712
301	E L S S P R E D T I P P S L T S C T E T S D P L P T N E N	1252206
183	E L A P P P - - - - - - - - - A C P S E S - - - - - -	GI 2660712
331	D D D I C K K P C S V A P N D I P L V S S T N L I N E I N G	1252206
195	- - - - - - - P V P T A P T A Q P - - - - - - - - - - -	GI 2660712
361	V S E K L S A T E S I V E I V K Q E V L P L T T E L E T I L E	1252206
205	- - - - - - - - - E E L L N G A P S P P A V D L S P V S	GI 2660712
391	N P P E E M K L E C I P A P I T P S T V P S F P P T P P T P	1252206
224	E P E E Q A K E - - V T A S V A P P T I P S A T P A P S	GI 2660712

FIGURE 2B

4/15

421	P A S P P H T P V I V P A A T T V S S P S A A I T V O R V	1252206
252	A T S P A Q E E E M E E E E E E E E E G E A G E A E S E	GI 2660712
451	L E E D E S I R T C L S E D A K E I O N K I E V E A D G Q T	1252206
282	K G G E E L - - - L P P E S T P I P A N L - - - - - - - -	GI 2660712
481	E E I L D S Q N L N S R R S P V V P A Q I A I T V P K T W K K	1252206
300	- - - S Q N L E A - - A A A T Q V A V S V P K R R R K	GI 2660712
511	P K D R T R T E M L E A E L E L K A E E L S I D K V L	1252206
322	I K E L N K K - - E A V G D L L D A F K E A N P A V P E V -	GI 2660712
541	E S E Q D K M S Q G F H P E R D P S D L K K V K A V E E N G	1252206
349	- - - E N Q P P A G S N P G P E S E G - - - S G V P P R P	GI 2660712
571	E E A E P V R N G A E S - V S E G E G I D A N S G S T D S S	1252206
372	E E A D E T W D S K E D K I H N A E N I Q P G E Q K - - -	GI 2660712
600	G D G V T F P F K P E S W K P T D T E G K K O Y D R E F L L	1252206
398	- - - - - Y E Y K S D Q W K P P N L E E K K R Y D R E F L L	GI 2660712

FIGURE 2C

5/15

630	D F Q F M P A C I Q K P E G L P P I S D V V L D K I N Q P K G F Q F I F A S M Q K P E G L P H I S D V V L D K A N - - K	1252206 GI 2660712
423		
660	L P M R T L D P R I I P R - - - G P D F T P A F A D F G R Q T P L R P L D P T R L Q G I N C G P D F T P S F A N L G R T	1252206 GI 2660712
451		
687	T P G G R G V P - - - - L L N V G S R R S Q T L S T R G P P R G G P G G E L P R G P Q A G L G P R R S Q	1252206 GI 2660712
481		
705	P G O R R E P R K I I - T V S V K E D V H L K K A E N A W K Q G P R K E P R K I I A T V L M T E D I K L N K A E K A W K	1252206 GI 2660712
511		
734	P S Q K R - - - - D S Q A D D P E N I K T Q E L F R K V R P S S K R T A A D K D R G E E D A D G S K T Q D L F R V R	1252206 GI 2660712
541		
759	S I L N K L T P Q M F N Q L M K Q V S G L T V D T E E R L K S I L N K L T P Q M F Q Q L M K Q V T Q L A T D T E E R L K	1252206 GI 2660712
571		
789	G V I D L V F E K A I D E P S F S V A Y A N M C R C L V T L G V I D L I F E K A I S E P N F S V A Y A N M C R C L M A L	1252206 GI 2660712
601		

FIGURE 2D

819 631	K V P M A D K P G N T V N F R K L L N R C Q K E F E K D K K V P T T E K P T V N F R K L L N R C Q K E F E K D K	1252206 GI 2660712
849 661	A D D V F E K K Q K E L E A A S A P E E R T R L H D E L E D D D E V F E K K Q K E M D E A A T A E E R G R L K E E L E	1252206 GI 2660712
879 691	E A K D K A R R R S I G N I K F I G E L F K L K M L T E A I E A R D T A R R R S I G N I K F I G E L F K L K M L T E A I	1252206 GI 2660712
909 721	M H D C V V K L L K N H D E E S L E C L C R L L T T I G K D M H D C V V K L L K N H D E E S L E C L C R L L T T I G K D	1252206 GI 2660712
939 751	L D F E K A K P R M D Q Y F N Q M E K I V K E K K T S S R I L D F E K A K P R M D Q Y F N Q M E K I T K E K K T S S R I	1252206 GI 2660712
969 781	R F M L Q D V I D L R L C N W V S R R A D Q G P K T I E Q I R F M L Q D V I D L R G S N W V P R R G D Q' G P K T I D Q I	1252206 GI 2660712
999 811	H K E A K I E E O E E O R K V Q Q L M T K - - E K R R - - - H K E A E M E E H R E H T K V Q Q L M A K G S D K R R G G P	1252206 GI 2660712

FIGURE 2E

7/15

1024	- - P G V Q R - - -	- - V D E G G W N T V Q G A K N S R V L D	1252206
841	P G P P I S R G L P L V D G G W N T V P I S K G S R P T D	GI 2660712	
1048	P S K F L K I T K P - T I D E K I Q L V P K A Q L G S W G K	1252206	
871	T S R L T K I T K P G S I D S N N Q L F A P G G R L S W G K	GI 2660712	
1077	G S S G G - - A K A S E T - - D A L R S S A S S L N R F S A	1252206	
901	G S S G G S G A K P S D A A S E A A R P A T S T L N R F S A	GI 2660712	
1103	L Q P P A P S G S T P S T P V E F D S R R T L T S R G S M G	1252206	
931	L Q Q A V P T E S T - - - - D N R R V V Q - R S S L S	GI 2660712	
1133	R E K N D K P L P S A T A R P N T F M R G G S S K D L L D N	1252206	
953	R E R G E K A - G D R G D R L E R S E R G G D R G D R L D R	GI 2660712	
1163	Q S Q E E Q R R - - - E M L E T V K O L T G G V D V E R N	1252206	
982	A R T P A T K R S F S K E V E E R S R E R P S Q P E G L R K	GI 2660712	
1189	S T E - - A E R N K T R E S A K P E I S A M S A H D - K A A	1252206	
1012	A A S L T E D R D R G R D A V K R E A A L P P V S P L K A A	GI 2660712	

FIGURE 2F

1216 L S E E E L E R K S K S I I D E F L H I N D F K E A M Q C V 1252206
 1042 L S E E E L E R K S K A I I E Y L H N D M K E A V Q C V GI 2660712

1246 E E L N A Q G L L H V F V R V G V E S T L E R S Q I T R D H 1252206
 1072 Q E L A S P S L L F I F V R H G V E S T L E R S A I A R E H GI 2660712

1276 M G Q L L Y Q L V O S E K L S K Q D F F K G F S E T I L E A 1252206
 1102 M G Q L L H Q L L C A G H L S T A Q Y Y Q G L Y E T I L E A GI 2660712

1306 D D M A I D I P H I W L Y L A E L V T P M L K E G G I S M R 1252206
 1132 E D M E I D I P H I V W L Y L A E L V T P T I L Q E G G V P M G GI 2660712

1336 E L T I E F S K P I L L P V G R A G V I L L S E I L H L C K Q 1252206
 1162 E L F R E I T K P I L R P I L G K A A S L L L E I L G L L C K S GI 2660712

1366 M S H K K V G A L W R E A D L S W K D F L P, E G E D V H N F 1252206
 1192 M G P K K V G T L W R E A G L S W K E F L P E G Q D I G A F GI 2660712

1396 L L E Q K L D F I E S D S P C S S E A L S K K E L S A E E L 1252206
 1222 V A E Q K V E Y T L G E - - - E S E A P G Q R A L P S E E L GI 2660712

FIGURE 2G

9/15

1426	Y K R L E K L I I E D K A N D E Q I F D W V E A N L D E I O	1252206
1249	N R Q L E K L L K E G - S S N Q R V F D W T E A N L S E Q Q	GI 2660712
1456	M S S P T F L R A L M T A V C K A A I I A D S S T F R V D T	1252206
1278	I V S N T L V R A L M T A V C Y S A I I F E T P - L R V D V	GI 2660712
1486	A V I K Q R V P I L L K Y L D S D T E K E L Q A L Y A L Q A	1252206
1307	A V L K A R A K L L Q K Y L - C D E Q K E L Q A L Y A L Q A	GI 2660712
1516	S I V K L D Q P A N L L R M F F D C L Y D E E V I S E D A F	1252206
1336	L V V T L E Q P P N L L R M F F D A L Y D E D V V K E D A F	GI 2660712
1546	Y K W E S S K D P A E Q N G K G V A L K S V T A F F T W L R	1252206
1366	Y S W E S S K D P A E Q Q G K G V A L K S V T A F F K W L R	GI 2660712
1576	E A E E S E D N	1252206
1396	E A E E S D H N	GI 2660712

FIGURE 2H

10/15

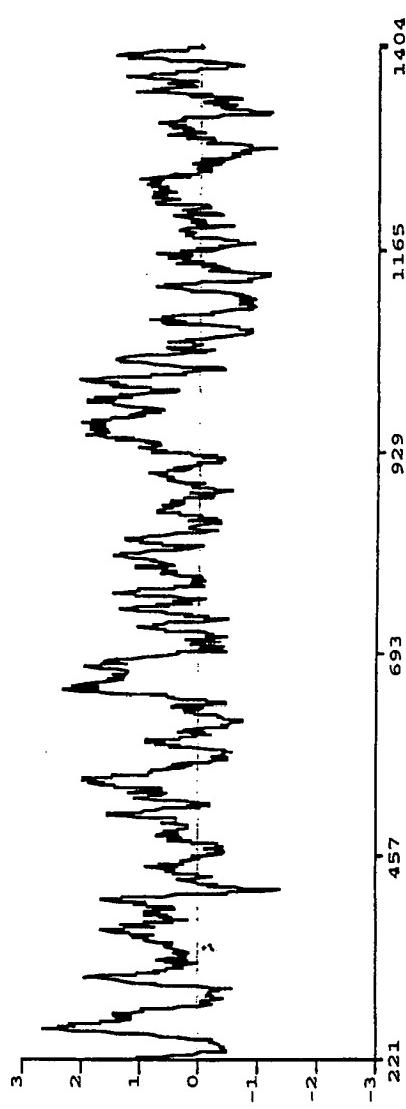


FIGURE 3A

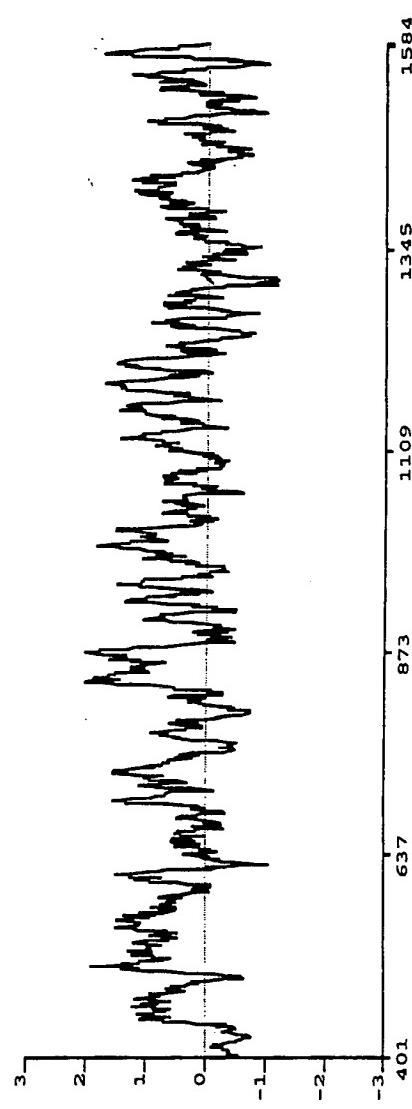


FIGURE 3B

11/15

1	M - - - - - L R R G A R N W R S M S T G E L T P Q S R L K E F S E L A R	2950994 GI 2440051
2	- - - - - A L N L Y R M D H L G N Y T G H K S V Y L T G Q L A T L E Q	2950994 GI 2440051
31	A I I Q Y A L Q A V T E H G F K L I S V P D I L P K E V I E	2950994 GI 2440051
2	- - - - - S C G M R T E G E R T Q V Y K L D T G E C L S G T S E M A L	2950994 GI 2440051
61	- - - - - G L - - - - -	F G V T G P 2950994
2	- - - - - A G F F A N K L L S E D Q L P L K V T A V S R C Y R A E T S	2950994 GI 2440051
91	- - - - - G L Q E E K G I Y R V H Q F N K V E M F A I C T E E Q S' S E A	E Q S S Q 2950994 GI 2440051
121	- - - - - G L - - - - -	G L - - - - -
8	- - - - - G L - - - - -	G L - - - - -
151	- - - - - G L - - - - -	G L - - - - -

FIGURE 4A

12/15

15 L L E E F L S L Q M E I L T E L G L H F R V L D M P T Q E L 2950994
 181 E L E E F K N I E V D L F R R L G L N F R L L D M P P C E L GI 2440051

45 G L P A Y R K F D I E A W M P G R G R F G E V T S A S N C T 2950994
 211 G A P A Y Q K Y D I E A W M P G R Q M W G E I S S C S N C T GI 2440051

75 D F Q S R R L H I M F Q T E A - G E L Q F A H T V N A T A C 2950994
 241 D Y Q A R R L G I R Y R R S A D G Q T I L H A H T I N G T A T GI 2440051

104 A V P R L L I A L L E S N Q Q K D G S V L V P P A L Q S Y L 2950994
 271 A T P R L L I A L L E S Y Q - K E D G I E I P A V L R P F M GI 2440051

134 G T D R - I T A P T H V P - - - L Q Y I G P N Q P R K P G 2950994
 300 D N Q E L I T R N K R I P E T K L V K F I K A GI 2440051

159 L P G Q P A V S
 322

FIGURE 4B

13/15

FIGURE 5A

14/15

181	E D L Q K M K Q F A Q D F V M H T D V R T C S S S T S V I A	3461657
25	- - - - -	GI 1808648
211	D L Q E D E D G V Y F S S Y G H Y G I H E E M L K D K I R T	3461657
25	- - - - - - - - - - -	GI 1808648
241	E S Y R D F I Y Q N P H I F K D K V V L D V G C G T G I L S	3461657
46	L T Y R N S M F H N R H L F K D K V V L D V G S G T G I L C	GI 1808648
271	M F A A K A G A K K V L G V D Q S E I L Y Q A M D I I R L N	3461657
76	M F A A K A G A R K V I G I V C S S I S D Y A V K I V K A N	GI 1808648
301	K L E D T I T L I K G K I E E V H L P V E K V D V I I S E W	3461657
106	K L D H V V T I K G K V E E V E L P V E K V D I I I S E W	GI 1808648
331	M G Y F L L F E S M L D S V L Y A K N K Y L A K G G S V Y P	3461657
136	M G Y C L F Y E S M L N T V L Y A R D K W L A P D G L I F P	GI 1808648

15/15

361	D I C T I S L V A V S D V N K H A D R I A F W D D V Y G F K	3461657
166	D R A T I L Y V T A T E D R Q Y K D Y K I H W W E N V Y G F D	GI 1808648
391	M S C M K K A V I P E A V V E V L D P K T L I S E P C G I K	3461657
196	M S C I K D V A I K E P L V D V V D P K Q L V T N A C L I K	GI 1808648
421	H I D C H T T S I S D L E F S S D F T L K I T R T S M C T A	3461657
226	E V D I Y T V K V E D L T F T S P F C L Q V K R N D Y V H A	GI 1808648
451	I A G Y F D I Y F E K N C H N R V V F S T G P Q S T K T H W	3461657
256	L V A Y F N I E F T R - C H K R T G F S T S P E S P Y T H W	GI 1808648
481	K Q T V F L L E K P F S V K A G E A L K G K V T V H K N K K	3461657
285	K Q T V F Y M E D Y L T V K T G E E I F G T I G M R P N A K	GI 1808648
511	D P R S L T V T L T L N - - - - - N S T Q T Y G L Q	3461657
315	N N R D L D F T I D L D F K G Q L C E L S C S T D Y R M R	GI 1808648

FIGURE 5C

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.

TANG, Y. Tom
CORLEY, Neil C.
GUEGLER, Karl J.
GORONE, Gina A.
PATTERSON, Chandra
HILLMAN, Jennifer L.
BAUGHN, Mariah R.
LAL, Preeti
AZIMZAI, Yalda
YUE, Henry
YANG, Junming

<120> RNA-ASSOCIATED PROTEINS

<130> PF-0600 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/156,039; unassigned; 09/158,720; unassigned; 09/186,815;
unassigned; 60/128,660

<151> 1998-09-17; 1998-09-17; 1998-09-22; 1998-09-22; 1998-11-04;
1998-11-04; 1999-04-08

<160> 38

<170> PERL Program

<210> 1

<211> 216

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 399781CD1

<400> 1

Met	Ser	Arg	Tyr	Leu	Arg	Pro	Pro	Asn	Thr	Ser	Leu	Phe	Val	Arg
1				5					10					15
Asn	Val	Ala	Asp	Asp	Thr	Arg	Ser	Glu	Asp	Leu	Arg	Arg	Glu	Phe
	"				20				25					30
Gly	Arg	Tyr	Gly	Pro	Ile	Val	Asp	Val	Tyr	Val	Pro	Leu	Asp	Phe
				35					40					45
Tyr	Thr	Arg	Arg	Pro	Arg	Gly	Phe	Ala	Tyr	Val	Gln	Phe	Glu	Asp
				50				55						60
Val	Arg	Asp	Ala	Glu	Asp	Ala	Leu	His	Asn	Leu	Asp	Arg	Lys	Trp
				65				70						75
Ile	Cys	Gly	Arg	Gln	Ile	Glu	Ile	Gln	Phe	Ala	Gln	Gly	Asp	Arg
				80				85						90
Lys	Thr	Pro	Asn	Gln	Met	Lys	Ala	Lys	Glu	Gly	Arg	Asn	Val	Tyr
				95				100						105
Ser	Ser	Ser	Arg	Tyr	Asp	Asp	Tyr	Asp	Arg	Tyr	Arg	Arg	Ser	Arg
				110				115						120

Ser Arg Ser Tyr Glu Arg Arg Arg Ser Arg Ser Arg Ser Phe Asp
 125 130 135
 Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro Arg Asn Ser Arg Pro Thr
 140 145 150
 Gly Arg Pro Arg Arg Arg Glu Ala Ile Pro Thr Met Ile Asp Gln
 155 160 165
 Thr Ala Ala Gly Ile Pro Ser Thr Val Leu Leu Thr Thr Leu Gln
 170 175 180
 Glu Arg Ser Glu Ser Gly Lys Arg Thr Lys Glu Gly Gln Phe Lys
 185 190 195
 Arg Pro Lys Gly Gly Trp Lys Val Leu Gln Tyr Glu Tyr Cys Thr
 200 205 210
 Asn Ile Leu Thr Leu Val
 215

<210> 2
 <211> 1584
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 1252206CD1

<400> 2
 Met Asn Ser Gln Pro Gln Thr Arg Ser Pro Phe Phe Gln Arg Pro
 1 5 10 15
 Gln Ile Gln Pro Pro Arg Ala Thr Ile Pro Asn Ser Ser Pro Ser
 20 25 30
 Ile Arg Pro Gly Ala Gln Thr Pro Thr Ala Val Tyr Gln Ala Asn
 35 40 45
 Gln His Ile Met Met Val Asn His Leu Pro Met Pro Tyr Pro Val
 50 55 60
Pro-Gln-Gly-Pro-Gln-Tyr-Cys-Ile_Pro_Gln_Tyr_Arg_His_Ser_Gly
 65 70 75
 Pro Pro Tyr Val Gly Pro Pro Gln Lys Tyr Pro Val Gln Pro Pro
 80 85 90
 Gly Pro Gly Pro Phe Tyr Pro Gly Pro Gly Pro Gly Asp Phe Pro
 95 100 105
 Asn Ala Tyr Gly Thr Pro Phe Tyr Pro Ser Gln Pro Val Tyr Gln
 110 115 120
 Ser Ala Pro Ile Ile Val Pro Thr Gln Gln Gln Pro Pro Pro Ala
 125 130 135
 Lys Arg Glu Lys Lys Thr Ile Arg Ile Arg Asp Pro Asn Gln Gly
 140 145 150
 Gly Lys Asp Ile Thr Glu Glu Ile Met Ser Gly Gly Ser Arg
 155 160 165
 Asn Pro Thr Pro Pro Ile Gly Arg Pro Thr Ser Thr Pro Thr Pro
 170 175 180
 Pro Gln Leu Pro Ser Gln Val Pro Glu His Ser Pro Val Val Tyr
 185 190 195
 Gly Thr Val Glu Ser Ala His Leu Ala Ala Ser Thr Pro Val Thr
 200 205 210
 Ala Ala Ser Asp Gln Lys Gln Glu Glu Lys Pro Lys Pro Asp Pro
 215 220 225
 Val Leu Lys Ser Pro Ser Pro Val Leu Arg Leu Val Leu Ser Gly

230	235	240
Glu Lys Lys Glu Gln	Glu Gly Gln Thr Ser	Glu Thr Thr Ala Ile
245	250	255
Val Ser Ile Ala Glu	Leu Pro Leu Pro Pro	Ser Pro Thr Thr Val
260	265	270
Ser Ser Val Ala Arg	Ser Thr Ile Ala Ala	Pro Thr Ser Ser Ala
275	280	285
Leu Ser Ser Gln Pro	Ile Phe Thr Thr Ala	Ile Asp Asp Arg Cys
290	295	300
Glu Leu Ser Ser Pro	Arg Glu Asp Thr Ile	Pro Ile Pro Ser Leu
305	310	315
Thr Ser Cys Thr Glu	Thr Ser Asp Pro Leu	Pro Thr Asn Glu Asn
320	325	330
Asp Asp Asp Ile Cys	Lys Lys Pro Cys Ser	Val Ala Pro Asn Asp
335	340	345
Ile Pro Leu Val Ser	Ser Ser Thr Asn Leu	Ile Asn Glu Ile Asn Gly
350	355	360
Val Ser Glu Lys Leu	Ser Ala Thr Glu Ser	Ile Val Glu Ile Val
365	370	375
Lys Gln Glu Val Leu	Pro Leu Thr Leu Glu	Leu Glu Ile Leu Glu
380	385	390
Asn Pro Pro Glu Glu	Met Lys Leu Glu Cys	Ile Pro Ala Pro Ile
395	400	405
Thr Pro Ser Thr Val	Pro Ser Phe Pro Pro	Thr Pro Pro Thr Pro
410	415	420
Pro Ala Ser Pro Pro	His Thr Pro Val Ile	Val Pro Ala Ala Ala
425	430	435
Thr Thr Val Ser Ser	Pro Ser Ala Ala Ile	Thr Val Gln Arg Val
440	445	450
Leu Glu Glu Asp Glu	Ser Ile Arg Thr Cys	Leu Ser Glu Asp Ala
455	460	465
Lys Glu Ile Gln Asn	Lys Ile Glu Val Glu	Ala Asp Gly Gln Thr
470	475	480
Glu Glu Ile Leu Asp	Ser Gln Asn Leu Asn	Ser Arg Arg Ser Pro
485	490	495
Val Pro Ala Gln Ile	Ala Ile Thr Val Pro	Lys Thr Trp Lys Lys
500	505	510
Pro Lys Asp Arg Thr	Arg Thr Thr Glu Glu	Met Leu Glu Ala Glu
515	520	525
Leu Glu Leu Lys Ala	Glu Glu Glu Leu Ser	Ile Asp Lys Val Leu
530	535	540
Glu Ser Glu Gln Asp	Lys Met Ser Gln Gly	Phe His Pro Glu Arg
545	550	555
Asp Pro Ser Asp Leu	Lys Val Lys Ala Val	Glu Asn Gly
560	565	570
Glu Glu Ala Glu Pro	Val Arg Asn Gly Ala	Glu Ser Val Ser Glu
575	580	585
Gly Glu Gly Ile Asp	Ala Asn Ser Gly Ser	Thr Asp Ser Ser Gly
590	595	600
Asp Gly Val Thr Phe	Pro Phe Lys Pro Glu	Ser Trp Lys Pro Thr
605	610	615
Asp Thr Glu Gly Lys	Lys Gln Tyr Asp Arg	Glu Phe Leu Leu Asp
620	625	630
Phe Gln Phe Met Pro	Ala Cys Ile Gln Lys	Pro Glu Gly Leu Pro
635	640	645
Pro Ile Ser Asp Val	Val Leu Asp Lys Ile	Asn Gln Pro Lys Leu
650	655	660

Pro Met Arg Thr Leu Asp Pro Arg Ile Leu Pro Arg Gly Pro Asp
 665 670 675
 Phe Thr Pro Ala Phe Ala Asp Phe Gly Arg Gln Thr Pro Gly Gly
 680 685 690
 Arg Gly Val Pro Leu Leu Asn Val Gly Ser Arg Arg Ser Gln Pro
 695 700 705
 Gly Gln Arg Arg Glu Pro Arg Lys Ile Ile Thr Val Ser Val Lys
 710 715 720
 Glu Asp Val His Leu Lys Lys Ala Glu Asn Ala Trp Lys Pro Ser
 725 730 735
 Gln Lys Arg Asp Ser Gln Ala Asp Asp Pro Glu Asn Ile Lys Thr
 740 745 750
 Gln Glu Leu Phe Arg Lys Val Arg Ser Ile Leu Asn Lys Leu Thr
 755 760 765
 Pro Gln Met Phe Asn Gln Leu Met Lys Gln Val Ser Gly Leu Thr
 770 775 780
 Val Asp Thr Glu Glu Arg Leu Lys Gly Val Ile Asp Leu Val Phe
 785 790 795
 Glu Lys Ala Ile Asp Glu Pro Ser Phe Ser Val Ala Tyr Ala Asn
 800 805 810
 Met Cys Arg Cys Leu Val Thr Leu Lys Val Pro Met Ala Asp Lys
 815 820 825
 Pro Gly Asn Thr Val Asn Phe Arg Lys Leu Leu Asn Arg Cys
 830 835 840
 Gln Lys Glu Phe Glu Lys Asp Lys Ala Asp Asp Asp Val Phe Glu
 845 850 855
 Lys Lys Gln Lys Glu Leu Glu Ala Ala Ser Ala Pro Glu Glu Arg
 860 865 870
 Thr Arg Leu His Asp Glu Leu Glu Ala Lys Asp Lys Ala Arg
 875 880 885
 Arg Arg Ser Ile Gly Asn Ile Lys Phe Ile Gly Glu Leu Phe Lys
 890 895 900
 Leu Lys Met Leu Thr Glu Ala Ile Met His Asp Cys Val Val Lys
 905 910 915
 Leu Leu Lys Asn His Asp Glu Glu Ser Leu Glu Cys Leu Cys Arg
 920 925 930
 Leu Leu Thr Thr Ile Gly Lys Asp Leu Asp Phe Glu Lys Ala Lys
 935 940 945
 Pro Arg Met Asp Gln Tyr Phe Asn Gln Met Glu Lys Ile Val Lys
 950 955 960
 Glu Lys Lys Thr Ser Ser Arg Ile Arg Phe Met Leu Gln Asp Val
 965 970 975
 Ile Asp Leu Arg Leu Cys Asn Trp Val Ser Arg Arg Ala Asp Gln
 980 985 990
 Gly Pro Lys Thr Ile Glu Gln Ile His Lys Glu Ala Lys Ile Glu
 995 1000 1005
 Glu Gln Glu Gln Arg Lys Val Gln Gln Leu Met Thr Lys Glu
 1010 1015 1020
 Lys Arg Arg Pro Gly Val Gln Arg Val Asp Glu Gly Gly Trp Asn
 1025 1030 1035
 Thr Val Gln Gly Ala Lys Asn Ser Arg Val Leu Asp Pro Ser Lys
 1040 1045 1050
 Phe Leu Lys Ile Thr Lys Pro Thr Ile Asp Glu Lys Ile Gln Leu
 1055 1060 1065
 Val Pro Lys Ala Gln Leu Gly Ser Trp Gly Lys Gly Ser Ser Gly
 1070 1075 1080
 Gly Ala Lys Ala Ser Glu Thr Asp Ala Leu Arg Ser Ser Ala Ser

1085	1090	1095
Ser Leu Asn Arg Phe Ser Ala Leu Gln Pro Pro Ala Pro Ser Gly		
1100	1105	1110
Ser Thr Pro Ser Thr Pro Val Glu Phe Asp Ser Arg Arg Thr Leu		
1115	1120	1125
Thr Ser Arg Gly Ser Met Gly Arg Glu Lys Asn Asp Lys Pro Leu		
1130	1135	1140
Pro Ser Ala Thr Ala Arg Pro Asn Thr Phe Met Arg Gly Gly Ser		
1145	1150	1155
Ser Lys Asp Leu Leu Asp Asn Gln Ser Gln Glu Glu Gln Arg Arg		
1160	1165	1170
Glu Met Leu Glu Thr Val Lys Gln Leu Thr Gly Gly Val Asp Val		
1175	1180	1185
Glu Arg Asn Ser Thr Glu Ala Glu Arg Asn Lys Thr Arg Glu Ser		
1190	1195	1200
Ala Lys Pro Glu Ile Ser Ala Met Ser Ala His Asp Lys Ala Ala		
1205	1210	1215
Leu Ser Glu Glu Leu Glu Arg Lys Ser Lys Ser Ile Ile Asp		
1220	1225	1230
Glu Phe Leu His Ile Asn Asp Phe Lys Glu Ala Met Gln Cys Val		
1235	1240	1245
Glu Glu Leu Asn Ala Gln Gly Leu Leu His Val Phe Val Arg Val		
1250	1255	1260
Gly Val Glu Ser Thr Leu Glu Arg Ser Gln Ile Thr Arg Asp His		
1265	1270	1275
Met Gly Gln Leu Leu Tyr Gln Leu Val Gln Ser Glu Lys Leu Ser		
1280	1285	1290
Lys Gln Asp Phe Phe Lys Gly Phe Ser Glu Thr Leu Glu Leu Ala		
1295	1300	1305
Asp Asp Met Ala Ile Asp Ile Pro His Ile Trp Leu Tyr Leu Ala		
1310	1315	1320
Glu Leu Val Thr Pro Met Leu Lys Glu Gly Gly Ile Ser Met Arg		
1325	1330	1335
Glu Leu Thr Ile Glu Phe Ser Lys Pro Leu Leu Pro Val Gly Arg		
1340	1345	1350
Ala Gly Val Leu Leu Ser Glu Ile Leu His Leu Leu Cys Lys Gln		
1355	1360	1365
Met Ser His Lys Lys Val Gly Ala Leu Trp Arg Glu Ala Asp Leu		
1370	1375	1380
Ser Trp Lys Asp Phe Leu Pro Glu Gly Glu Asp Val His Asn Phe		
1385	1390	1395
Leu Leu Glu Gln Lys Leu Asp Phe Ile Glu Ser Asp Ser Pro Cys		
1400	1405	1410
Ser Ser Glu Ala Leu Ser Lys Lys Glu Leu Ser Ala Glu Glu Leu		
1415	1420	1425
Tyr Lys Arg Leu Glu Lys Leu Ile Ile Glu Asp Lys Ala Asn Asp		
1430	1435	1440
Glu Gln Ile Phe Asp Trp Val Glu Ala Asn Leu Asp Glu Ile Gln		
1445	1450	1455
Met Ser Ser Pro Thr Phe Leu Arg Ala Leu Met Thr Ala Val Cys		
1460	1465	1470
Lys Ala Ala Ile Ile Ala Asp Ser Ser Thr Phe Arg Val Asp Thr		
1475	1480	1485
Ala Val Ile Lys Gln Arg Val Pro Ile Leu Leu Lys Tyr Leu Asp		
1490	1495	1500
Ser Asp Thr Glu Lys Glu Leu Gln Ala Leu Tyr Ala Leu Gln Ala		
1505	1510	1515

Ser Ile Val Lys Leu Asp Gln Pro Ala Asn Leu Leu Arg Met Phe
 1520 1525 1530
 Phe Asp Cys Leu Tyr Asp Glu Glu Val Ile Ser Glu Asp Ala Phe
 1535 1540 1545
 Tyr Lys Trp Glu Ser Ser Lys Asp Pro Ala Glu Gln Asn Gly Lys
 1550 1555 1560
 Gly Val Ala Leu Lys Ser Val Thr Ala Phe Phe Thr Trp Leu Arg
 1565 1570 1575
 Glu Ala Glu Glu Ser Glu Asp Asn
 1580

<210> 3
 <211> 166
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2950994CD1

<400> 3
 Met Phe Gly Val Thr Gly Pro Gly Leu Glu Gln Ser Ser Gln Leu
 1 5 10 15
 Leu Glu Glu Phe Leu Ser Leu Gln Met Glu Ile Leu Thr Glu Leu
 20 25 30
 Gly Leu His Phe Arg Val Leu Asp Met Pro Thr Gln Glu Leu Gly
 35 40 45
 Leu Pro Ala Tyr Arg Lys Phe Asp Ile Glu Ala Trp Met Pro Gly
 50 55 60
 Arg Gly Arg Phe Gly Glu Val Thr Ser Ala Ser Asn Cys Thr Asp
 65 70 75
 Phe Gln Ser Arg Arg Leu His Ile Met Phe Gln Thr Glu Ala Gly
 80 85 90
Glu_Leu_Gln_Phe_Ala_His_Thr_Val_Asn_Ala_Thr_Ala_Cys_Ala_Val
 95 100 105
 Pro Arg Leu Leu Ile Ala Leu Leu Glu Ser Asn Gln Gln Lys Asp
 110 115 120
 Gly Ser Val Leu Val Pro Pro Ala Leu Gln Ser Tyr Leu Gly Thr
 125 130 135
 Asp Arg Ile Thr Ala Pro Thr His Val Pro Leu Gln Tyr Ile Gly
 140 145 150
 Pro Asn Gln Pro Arg Lys Pro Gly Leu Pro Gly Gln Pro Ala Val
 155 160 165

Ser

<210> 4
 <211> 531
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3461657CD1

<400> 4

Met Cys Ser Leu Ala Ser Gly Ala Thr Gly Gly Arg Gly Ala Val
 1 5 10 15
 Glu Asn Glu Glu Asp Leu Pro Glu Leu Ser Asp Ser Gly Asp Glu
 20 25 30
 Ala Ala Trp Glu Asp Glu Asp Ala Asp Leu Pro His Gly Lys
 35 40 45
 Gln Gln Thr Pro Cys Leu Phe Cys Asn Arg Leu Phe Thr Ser Ala
 50 55 60
 Glu Glu Thr Phe Ser His Cys Lys Ser Glu His Gln Phe Asn Ile
 65 70 75
 Asp Ser Met Val His Lys His Gly Leu Glu Phe Tyr Gly Tyr Ile
 80 85 90
 Lys Leu Ile Asn Phe Ile Arg Leu Lys Asn Pro Thr Val Glu Tyr
 95 100 105
 Met Asn Ser Ile Tyr Asn Pro Val Pro Trp Glu Lys Glu Glu Tyr
 110 115 120
 Leu Lys Pro Val Leu Glu Asp Asp Leu Leu Gln Phe Asp Val
 125 130 135
 Glu Asp Leu Tyr Glu Pro Val Ser Val Pro Phe Ser Tyr Pro Asn
 140 145 150
 Gly Leu Ser Glu Asn Thr Ser Val Val Glu Lys Leu Lys His Met
 155 160 165
 Glu Ala Arg Ala Leu Ser Ala Glu Ala Ala Leu Ala Arg Ala Arg
 170 175 180
 Glu Asp Leu Gln Lys Met Lys Gln Phe Ala Gln Asp Phe Val Met
 185 190 195
 His Thr Asp Val Arg Thr Cys Ser Ser Ser Thr Ser Val Ile Ala
 200 205 210
 Asp Leu Gln Glu Asp Glu Asp Gly Val Tyr Phe Ser Ser Tyr Gly
 215 220 225
 His Tyr Gly Ile His Glu Glu Met Leu Lys Asp Lys Ile Arg Thr
 230 235 240
 Glu Ser Tyr Arg Asp Phe Ile Tyr Gln Asn Pro His Ile Phe Lys
 245 250 255
 Asp Lys Val Val Leu Asp Val Gly Cys Gly Thr Gly Ile Leu Ser
 260 265 270
 Met Phe Ala Ala Lys Ala Gly Ala Lys Lys Val Leu Gly Val Asp
 275 280 285
 Gln Ser Glu Ile Leu Tyr Gln Ala Met Asp Ile Ile Arg Leu Asn
 290 295 300
 Lys Leu Glu Asp Thr Ile Thr Leu Ile Lys Gly Lys Ile Glu Glu
 305 310 315
 Val His Leu Pro Val Glu Lys Val Asp Val Ile Ile Ser Glu Trp
 320 325 330
 Met Gly Tyr Phè Leu Leu Phe Glu Ser Met Leu Asp Ser Val Leu
 335 340 345
 Tyr Ala Lys Asn Lys Tyr Leu Ala Lys Gly Gly Ser Val Tyr Pro
 350 355 360
 Asp Ile Cys Thr Ile Ser Leu Val Ala Val Ser Asp Val Asn Lys
 365 370 375
 His Ala Asp Arg Ile Ala Phe Trp Asp Asp Val Tyr Gly Phe Lys
 380 385 390
 Met Ser Cys Met Lys Lys Ala Val Ile Pro Glu Ala Val Val Glu
 395 400 405
 Val Leu Asp Pro Lys Thr Leu Ile Ser Glu Pro Cys Gly Ile Lys
 410 415 420
 His Ile Asp Cys His Thr Thr Ser Ile Ser Asp Leu Glu Phe Ser

425	430	435
Ser Asp Phe Thr Leu Lys Ile Thr Arg	Thr Ser Met Cys Thr Ala	
440	445	450
Ile Ala Gly Tyr Phe Asp Ile Tyr Phe	Glu Lys Asn Cys His Asn	
455	460	465
Arg Val Val Phe Ser Thr Gly Pro Gln	Ser Thr Lys Thr His Trp	
470	475	480
Lys Gln Thr Val Phe Leu Leu Glu Lys	Pro Phe Ser Val Lys Ala	
485	490	495
Gly Glu Ala Leu Lys Gly Lys Val Thr	Val His Lys Asn Lys Lys	
500	505	510
Asp Pro Arg Ser Leu Thr Val Thr Leu	Thr Leu Asn Asn Ser Thr	
515	520	525
Gln Thr Tyr Gly Leu Gln		
530		

<210> 5
<211> 148
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 053076CD1

<400> 5		
Met Ala Ser Val Val Leu Ala Leu Arg	Thr Arg Thr Ala Val Thr	
1	5	10
Ser Leu Leu Ser Pro Thr Pro Ala Thr	Ala Leu Ala Val Arg Tyr	
20	25	30
Ala Ser Lys Lys Ser Gly Gly Ser Ser	Lys Asn Leu Gly Gly Lys	
35	40	45
Ser Ser Gly Arg Arg Gln Gly Ile Lys	Lys Met Glu Gly His Tyr	
50	55	60
Val His Ala Gly Asn Ile Ile Ala Thr	Gln Arg His Phe Arg Trp	
65	70	75
His Pro Gly Ala His Val Gly Val Gly	Lys Asn Lys Cys Leu Tyr	
80	85	90
Ala Leu Glu Glu Gly Ile Val Arg Tyr	Thr Lys Glu Val Tyr Val	
95	100	105
Pro His Pro Arg Asn Thr Glu Ala Val	Asp Leu Ile Thr Arg Leu	
110	115	120
Pro Lys Gly Ala Val Leu Tyr Lys Thr	Phe Val His Val Val Pro	
125	130	135
Ala Lys Pro Glu Gly Thr Phe Lys Leu	Val Ala Met Leu	
140	145	

<210> 6
<211> 317
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1292379CD1

<400> 6

Met	Met	Ser	Phe	His	Ser	Asn	Arg	Pro	Ser	Lys	Arg	Phe	Cys	Ile
1					5				10					15
Phe	Lys	Lys	His	Ser	Glu	Asn	Leu	Arg	Gly	Ile	Thr	Leu	Val	Cys
					20				25					30
Leu	Asn	Cys	Asp	Phe	Leu	Ser	Asp	Val	Ser	Gly	Leu	Asp	Asn	Met
					35				40					45
Ala	Thr	His	Leu	Ser	Gln	His	Lys	Thr	His	Thr	Cys	Gln	Val	Val
					50				55					60
Met	Gln	Lys	Val	Ser	Val	Cys	Ile	Pro	Thr	Ser	Glu	His	Leu	Ser
					65				70					75
Glu	Leu	Lys	Lys	Glu	Ala	Pro	Ala	Lys	Glu	Gln	Glu	Pro	Val	Ser
					80				85					90
Lys	Glu	Ile	Ala	Arg	Pro	Asn	Met	Ala	Glu	Arg	Glu	Thr	Glu	Thr
					95				100					105
Ser	Asn	Ser	Glu	Ser	Lys	Gln	Asp	Lys	Ala	Ala	Ser	Ser	Lys	Glu
					110				115					120
Lys	Asn	Gly	Cys	Asn	Ala	Asn	Ser	Phe	Glu	Gly	Ser	Ser	Thr	Thr
					125				130					135
Lys	Ser	Glu	Glu	Ser	Ile	Thr	Val	Ser	Asp	Lys	Glu	Asn	Glu	Thr
					140				145					150
Cys	Leu	Ala	Asp	Gln	Glu	Thr	Gly	Ser	Lys	Asn	Ile	Val	Ser	Cys
					155				160					165
Asp	Ser	Asn	Ile	Gly	Ala	Asp	Lys	Val	Glu	Lys	Lys	Lys	Gln	Ile
					170				175					180
Gln	His	Val	Cys	Gln	Glu	Met	Glu	Leu	Lys	Met	Cys	Gln	Ser	Ser
					185				190					195
Glu	Asn	Ile	Ile	Leu	Ser	Asp	Gln	Ile	Lys	Asp	His	Asn	Ser	Ser
					200				205					210
Glu	Ala	Arg	Phe	Ser	Ser	Lys	Asn	Ile	Lys	Asp	Leu	Arg	Leu	Ala
					215				220					225
Ser	Asp	Asn	Val	Ser	Ile	Asp	Gln	Phe	Leu	Arg	Lys	Arg	His	Glu
					230				235					240
Pro	Glu	Ser	Val	Ser	Ser	Asp	Val	Ser	Glu	Gln	Gly	Ser	Ile	His
					245				250					255
Leu	Glu	Pro	Leu	Thr	Pro	Ser	Glu	Val	Leu	Glu	Tyr	Glu	Ala	Thr
					260				265					270
Glu	Ile	Leu	Gln	Lys	Gly	Ser	Gly	Asp	Pro	Ser	Ala	Lys	Thr	Asp
					275				280					285
Glu	Val	Val	Ser	Asp	Gln	Thr	Asp	Asp	Ile	Pro	Gly	Gly	Asn	Asn
					290				295					300
Pro	Ser	Thr	Thr	Glu	Ala	Thr	Val	Asp	Leu	Glu	Asp	Glu	Lys	Glu
					305				310					315
Arg	Ser													

<210> 7

<211> 278

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 1437783CD1

<400> 7

Met Ala Ala Leu Phe Leu Lys Arg Leu Thr Leu Gln Thr Val Lys		
1	5	10
Ser Glu Asn Ser Cys Ile Arg Cys Phe Gly Lys His Ile Leu Gln		
20	25	30
Lys Thr Ala Pro Ala Gln Leu Ser Pro Ile Ala Ser Ala Pro Arg		
35	40	45
Leu Ser Phe Leu Ile His Ala Lys Ala Phe Ser Thr Ala Glu Asp		
50	55	60
Thr Gln Asn Glu Gly Lys Lys Thr Lys Lys Asn Lys Thr Ala Phe		
65	70	75
Ser Asn Val Gly Arg Lys Ile Ser Gln Arg Val Ile His Leu Phe		
80	85	90
Asp Glu Lys Gly Asn Asp Leu Gly Asn Met His Arg Ala Asn Val		
95	100	105
Ile Arg Leu Met Asp Glu Arg Asp Leu Arg Leu Val Gln Arg Asn		
110	115	120
Thr Ser Thr Glu Pro Ala Glu Tyr Gln Leu Met Thr Gly Leu Gln		
125	130	135
Ile Leu Gln Glu Arg Gln Arg Leu Arg Glu Met Glu Lys Ala Asn		
140	145	150
Pro Lys Thr Gly Pro Thr Leu Arg Lys Glu Leu Ile Leu Ser Ser		
155	160	165
Asn Ile Gly Gln His Asp Leu Asp Thr Lys Thr Lys Gln Ile Gln		
170	175	180
Gln Trp Ile Lys Lys His Leu Val Gln Ile Thr Ile Lys Lys		
185	190	195
Gly Lys Asn Val Asp Val Ser Glu Asn Glu Met Glu Glu Ile Phe		
200	205	210
His Gln Ile Leu Gln Thr Met Pro Gly Ile Ala Thr Phe Ser Ser		
215	220	225
Arg Pro Gln Ala Val Gln Gly Gly Lys Ala Leu Met Cys Val Leu		
230	235	240
Arg Ala Leu Ser Lys Asn Glu Glu Lys Ala Tyr Lys Glu Thr Gln		
245	250	255
Glu Thr Gln Glu Arg Asp Thr Leu Asn Lys Asp His Gly Asn Asp		
260	265	270
Lys Glu Ser Asn Val Leu His Gln		
275		

<210> 8

<211> 586

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 1557635CD1

<400> 8

Met Ser Ala Thr Val Val Asp Ala Val Asn Ala Ala Pro Leu Ser		
1	5	10
Gly Ser Lys Glu Met Ser Leu Glu Glu Pro Lys Lys Met Thr Arg		
20	25	30
Glu Asp Trp Arg Lys Lys Glu Leu Glu Glu Gln Arg Lys Leu		
35	40	45

Gly Asn Ala Pro Ala Glu Val Asp Glu Glu Gly Lys Asp Ile Asn
 50 55 60
 Pro His Ile Pro Gln Tyr Ile Ser Ser Val Pro Trp Tyr Ile Asp
 65 70 75
 Pro Ser Lys Arg Pro Thr Leu Lys His Gln Arg Pro Gln Pro Glu
 80 85 90
 Lys Gln Lys Gln Phe Ser Ser Ser Gly Glu Trp Tyr Lys Arg Gly
 95 100 105
 Val Lys Glu Asn Ser Ile Ile Thr Lys Tyr Arg Lys Gly Ala Cys
 110 115 120
 Glu Asn Cys Gly Ala Met Thr His Lys Lys Lys Asp Cys Phe Glu
 125 130 135
 Arg Pro Arg Arg Val Gly Ala Lys Phe Thr Gly Thr Asn Ile Ala
 140 145 150
 Pro Asp Glu His Val Gln Pro Gln Leu Met Phe Asp Tyr Asp Gly
 155 160 165
 Lys Arg Asp Arg Trp Asn Gly Tyr Asn Pro Glu Glu His Met Lys
 170 175 180
 Ile Val Glu Glu Tyr Ala Lys Val Asp Leu Ala Lys Arg Thr Leu
 185 190 195
 Lys Ala Gln Lys Leu Gln Glu Glu Leu Ala Ser Gly Lys Leu Val
 200 205 210
 Glu Gln Ala Asn Ser Pro Lys His Gln Trp Gly Glu Glu Glu Pro
 215 220 225
 Asn Ser Gln Thr Glu Lys Asp His Asn Ser Glu Asp Glu Asp Glu
 230 235 240
 Asp Lys Tyr Ala Asp Asp Ile Asp Met Pro Gly Gln Asn Phe Asp
 245 250 255
 Ser Lys Arg Arg Ile Thr Val Arg Asn Leu Arg Ile Arg Glu Asp
 260 265 270
 Ile Ala Lys Tyr Leu Arg Asn Leu Asp Pro Asn Ser Ala Tyr Tyr
 275 280 285
 Asp Pro Lys Thr Arg Ala Met Arg Glu Asn Pro Tyr Ala Asn Ala
 290 295 300
 Gly Lys Asn Pro Asp Glu Val Ser Tyr Ala Gly Asp Asn Phe Val
 305 310 315
 Arg Tyr Thr Gly Asp Thr Ile Ser Met Ala Gln Thr Gln Leu Phe
 320 325 330
 Ala Trp Glu Ala Tyr Asp Lys Gly Ser Glu Val His Leu Gln Ala
 335 340 345
 Asp Pro Thr Lys Leu Glu Leu Leu Tyr Lys Ser Phe Lys Val Lys
 350 355 360
 Lys Glu Asp Phe Lys Glu Gln Gln Lys Glu Ser Ile Leu Glu Lys
 365 370 375
 Tyr Gly Gly Gln Glu His Leu Asp Ala Pro Pro Ala Glu Leu Leu
 380 385 390
 Leu Ala Gln Thr Glu Asp Tyr Val Glu Tyr Ser Arg His Gly Thr
 395 400 405
 Val Ile Lys Gly Gln Glu Arg Ala Val Ala Cys Ser Lys Tyr Glu
 410 415 420
 Glu Asp Val Lys Ile His Asn His Thr His Ile Trp Gly Ser Tyr
 425 430 435
 Trp Lys Glu Gly Arg Trp Gly Tyr Lys Cys Cys His Ser Phe Phe
 440 445 450
 Lys Tyr Ser Tyr Cys Thr Gly Glu Ala Gly Lys Glu Ile Val Asn
 455 460 465
 Ser Glu Glu Cys Ile Ile Asn Glu Ile Thr Gly Glu Glu Ser Val

WO 00/15799

470	475	480
Lys Lys Pro Gln Thr Leu Met Glu Leu His Gln Glu Lys Leu Lys		
485	490	495
Glu Glu Lys Lys Lys Lys Lys Lys Lys Lys His Arg Lys		
500	505	510
Ser Ser Ser Asp Ser Asp Asp Glu Glu Lys Lys His Glu Lys Leu		
515	520	525
Lys Lys Ala Leu Asn Ala Glu Glu Ala Arg Leu Leu His Val Lys		
530	535	540
Glu Thr Met Gln Ile Asp Glu Arg Lys Arg Pro Tyr Asn Ser Met		
545	550	555
Tyr Glu Thr Arg Glu Pro Thr Glu Glu Glu Met Glu Ala Tyr Arg		
560	565	570
Met Lys Arg Gln Arg Pro Asp Asp Pro Met Ala Ser Phe Leu Gly		
575	580	585

Gln

<210> 9
<211> 384
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2049352CD1

<400> 9

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp	15
1 5 10	
Thr Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg	30
20 25	
Asn Thr Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser	45
35 40	
Arg Pro Pro Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser	60
50 55	
Glu Leu Lys Tyr His Pro Glu Met Arg Phe Phe His Trp Phe Ser	75
65 70	
Lys Trp Arg Lys Leu His Arg Asp Gln Glu Tyr Glu Val Thr Trp	90
80 85	
Tyr Ile Ser Trp Ser Pro Cys Thr Lys Cys Thr Arg Asp Met Ala	105
95 100	
Thr Phe Leu Ala Glu Asp Pro Lys Val Thr Leu Thr Ile Phe Val	120
110 115	
Ala Arg Leu Tyr Tyr Phe Trp Asp Pro Asp Tyr Gln Glu Ala Leu	135
125 130	
Arg Ser Leu Cys Gln Lys Arg Asp Gly Pro Arg Ala Thr Met Lys	150
140 145	
Ile Met Asn Tyr Asp Glu Phe Gln His Cys Trp Ser Lys Phe Val	165
155 160	
Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn Leu Pro Lys	180
170 175	
Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile Leu Arg His	195
185 190	
Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro	210
200 205	
Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu	

215	220	225
Arg Met His Asn Asp Thr Trp Val Leu	Leu Asn Gln Arg Arg	Gly
230	235	240
Phe Leu Cys Asn Gln Ala Pro His Lys	His Gly Phe Leu Glu	Gly
245	250	255
Arg His Ala Glu Leu Cys Phe Leu Asp	Val Ile Pro Phe Trp	Lys
260	265	270
Leu Asp Leu Asp Gln Asp Tyr Arg Val	Thr Cys Phe Thr Ser	Trp
275	280	285
Ser Pro Cys Phe Ser Cys Ala Gln Glu	Met Ala Lys Phe Ile	Ser
290	295	300
Lys Asn Lys His Val Ser Leu Cys Ile	Phe Thr Ala Arg Ile	Tyr
305	310	315
Asp Asp Gln Gly Arg Cys Gln Glu Gly	Leu Arg Thr Leu Ala	Glu
320	325	330
Ala Gly Ala Lys Ile Ser Ile Leu Thr	Tyr Ser Glu Phe Lys	His
335	340	345
Cys Trp Asp Thr Phe Val Asp His Gln	Gly Cys Pro Phe Gln	Pro
350	355	360
Trp Asp Gly Leu Glu Glu His Ser Gln	Ala Leu Ser Gly Arg	Leu
365	370	375
Arg Gly Ile Leu Gln Asn Gln Gly Ser		
380		

<210> 10
<211> 325
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2231663CD1

<400> 10			
Met Ala Ala Ala Val Arg Cys Met Gly Arg Ala Leu Ile His His			
1	5	10	15
Gln Arg His Ser Leu Ser Lys Met Val Tyr Gln Thr Ser Leu Cys			
20	25	30	
Ser Cys Ser Val Asn Ile Arg Val Pro Asn Arg His Phe Ala Ala			
35	40	45	
Ala Thr Lys Ser Ala Lys Lys Thr Lys Lys Gly Ala Lys Glu Lys			
50	55	60	
Thr Pro Asp Glu Lys Lys Asp Glu Ile Glu Lys Ile Lys Ala Tyr			
65	70	75	
Pro Tyr Met Glu Gly Glu Pro Glu Asp Asp Val Tyr Leu Lys Arg			
80	85	90	
Leu Tyr Pro Arg Gln Ile Tyr Glu Val Glu Lys Ala Val His Leu			
95	100	105	
Leu Lys Lys Phe Gln Ile Leu Asp Phe Thr Ser Pro Lys Gln Ser			
110	115	120	
Val Tyr Leu Asp Leu Thr Leu Asp Met Ala Leu Gly Lys Lys Lys			
125	130	135	
Asn Val Glu Pro Phe Thr Ser Val Leu Ser Leu Pro Tyr Pro Phe			
140	145	150	
Ala Ser Glu Ile Asn Lys Val Ala Val Phe Thr Glu Asn Ala Ser			
155	160	165	

WO 00/15799

Glu Val Lys Ile Ala Glu Glu Asn Gly Ala Ala Phe Ala Gly Gly
 170 175 180
 Thr Ser Leu Ile Gln Lys Ile Trp Asp Asp Glu Ile Val Ala Asp
 185 190 195
 Phe Tyr Val Ala Val Pro Glu Ile Met Pro Glu Leu Asn Arg Leu
 200 205 210
 Arg Lys Lys Leu Asn Lys Tyr Pro Lys Leu Ser Arg Asn Ser
 215 220 225
 Ile Gly Arg Asp Ile Pro Lys Met Leu Glu Leu Phe Lys Asn Gly
 230 235 240
 His Glu Ile Lys Val Asp Glu Glu Arg Glu Asn Phe Leu Gln Thr
 245 250 255
 Lys Ile Ala Thr Leu Asp Met Ser Ser Asp Gln Ile Ala Ala Asn
 260 265 270
 Leu Gln Ala Val Ile Asn Glu Val Cys Arg His Arg Pro Leu Asn
 275 280 285
 Leu Gly Pro Phe Val Val Arg Ala Phe Leu Arg Ser Ser Thr Ser
 290 295 300
 Glu Gly Leu Leu Leu Lys Ile Asp Pro Leu Leu Pro Lys Glu Val
 305 310 315
 Lys Asn Glu Glu Ser Glu Lys Glu Asp Ala
 320 325

<210> 11
 <211> 351
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2604449CD1

<400> 11
 Met-Gly-Asp-Pro-Glu-Arg-Pro-Glu-Ala-Ala-Gly-Leu-Asp-Gln-Asp
 1 5 10 15
 Glu Arg Ser Ser Ser Asp Thr Asn Glu Ser Glu Ile Lys Ser Asn
 20 25 30
 Glu Glu Pro Leu Leu Arg Lys Ser Ser Arg Arg Phe Val Ile Phe
 35 40 45
 Pro Ile Gln Tyr Pro Asp Ile Trp Lys Met Tyr Lys Gln Ala Gln
 50 55 60
 Ala Ser Phe Trp Thr Ala Glu Glu Val Asp Leu Ser Lys Asp Leu
 65 70 75
 Pro His Trp Asn Lys Leu Lys Ala Asp Glu Lys Tyr Phe Ile Ser
 80 85 90
 His Ile Leu Ala Phe Phe Ala Ala Ser Asp Gly Ile Val Asn Glu
 95 100 105
 Asn Leu Val Glu Arg Phe Ser Gln Glu Val Gln Val Pro Glu Ala
 110 115 120
 Arg Cys Phe Tyr Gly Phe Gln Ile Leu Ile Glu Asn Val His Ser
 125 130 135
 Glu Met Tyr Ser Leu Leu Ile Asp Thr Tyr Ile Arg Asp Pro Lys
 140 145 150
 Lys Arg Glu Phe Leu Phe Asn Ala Ile Glu Thr Met Pro Tyr Val
 155 160 165
 Lys Lys Lys Ala Asp Trp Ala Leu Arg Trp Ile Ala Asp Arg Lys

	170	175	180
Ser Thr Phe Gly	Glu Arg Val Val Ala	Phe Ala Ala Val Glu	Gly
	185	190	195
Val Phe Phe Ser	Gly Ser Phe Ala Ala	Ile Phe Trp Leu Lys	Lys
	200	205	210
Arg Gly Leu Met	Pro Gly Leu Thr Phe	Ser Asn Glu Leu Ile	Ser
	215	220	225
Arg Asp Glu Gly	Leu His Cys Asp Phe	Ala Cys Leu Met Phe	Gln
	230	235	240
Tyr Leu Val Asn	Lys Pro Ser Glu Glu	Arg Val Arg Glu Ile	Ile
	245	250	255
Val Asp Ala Val	Lys Ile Glu Gln Glu	Phe Leu Thr Glu Ala	Leu
	260	265	270
Pro Val Gly Leu	Ile Gly Met Asn Cys	Ile Leu Met Lys Gln	Tyr
	275	280	285
Ile Glu Phe Val	Ala Asp Arg Leu Leu	Val Glu Leu Gly Phe	Ser
	290	295	300
Lys Val Phe Gln	Ala Glu Asn Pro Phe	Asp Phe Met Glu Asn	Ile
	305	310	315
Ser Leu Glu Gly	Lys Thr Asn Phe Phe	Glu Lys Arg Val Ser	Glu
	320	325	330
Tyr Gln Arg Phe	Ala Val Met Ala Glu	Thr Thr Asp Asn Val	Phe
	335	340	345
Thr Leu Asp Ala	Asp Phe		
	350		

<210> 12
<211> 681
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2604993CD1

	<400> 12		
Met Thr Ala Ser Pro Asp Tyr Leu Val Val Leu Phe Gly Ile Thr			
1	5	10	15
Ala Gly Ala Thr Gly Ala Lys Leu Gly Ser Asp Glu Lys Glu Leu			
	20	25	30
Ile Leu Leu Phe Trp Lys Val Val Asp Leu Ala Asn Lys Lys Val			
	35	40	45
Gly Gln Leu His Glu Val Leu Val Arg Pro Asp Gln Leu Glu Leu			
	50	55	60
Thr Glu Asp Cys Lys Glu Glu Thr Lys Ile Asp Val Glu Ser Leu			
	65	70	75
Ser Ser Ala Ser Gln Leu Asp Gln Ala Leu Arg Gln Phe Asn Gln			
	80	85	90
Ser Val Ser Asn Glu Leu Asn Ile Gly Val Gly Thr Ser Phe Cys			
	95	100	105
Leu Cys Thr Asp Gly Gln Leu His Val Arg Gln Ile Leu His Pro			
	110	115	120
Glu Ala Ser Lys Lys Asn Val Leu Leu Pro Glu Cys Phe Tyr Ser			
	125	130	135
Phe Phe Asp Leu Arg Lys Glu Phe Lys Lys Cys Cys Pro Gly Ser			
	140	145	150

Pro Asp Ile Asp Lys Leu Asp Val Ala Thr Met Thr Glu Tyr Leu
 155 160 165
 Asn Phe Glu Lys Ser Ser Ser Val Ser Arg Tyr Gly Ala Ser Gln
 170 175 180
 Val Glu Asp Met Gly Asn Ile Ile Leu Ala Met Ile Ser Glu Pro
 185 190 195
 Tyr Asn His Arg Phe Ser Asp Pro Glu Arg Val Asn Tyr Lys Phe
 200 205 210
 Glu Ser Gly Thr Cys Ser Lys Met Glu Leu Ile Asp Asp Asn Thr
 215 220 225
 Val Val Arg Ala Arg Gly Leu Pro Trp Gln Ser Ser Asp Gln Asp
 230 235 240
 Ile Ala Arg Phe Phe Lys Gly Leu Asn Ile Ala Lys Gly Gly Ala
 245 250 255
 Ala Leu Cys Leu Asn Ala Gln Gly Arg Arg Asn Gly Glu Ala Leu
 260 265 270
 Val Arg Phe Val Ser Glu Glu His Arg Asp Leu Ala Leu Gln Arg
 275 280 285
 His Lys His His Met Gly Thr Arg Tyr Ile Glu Val Tyr Lys Ala
 290 295 300
 Thr Gly Glu Asp Phe Leu Lys Ile Ala Gly Gly Thr Ser Asn Glu
 305 310 315
 Val Ala Gln Phe Leu Ser Lys Glu Asn Gln Val Ile Val Arg Met
 320 325 330
 Arg Gly Leu Pro Phe Thr Ala Thr Ala Glu Glu Val Val Ala Phe
 335 340 345
 Phe Gly Gln His Cys Pro Ile Thr Gly Gly Lys Glu Gly Ile Leu
 350 355 360
 Phe Val Thr Tyr Pro Asp Gly Arg Pro Thr Gly Asp Ala Phe Val
 365 370 375
 Leu Phe Ala Cys Glu Glu Tyr Ala Gln Asn Ala Leu Arg Lys His
 380 385 390
 Lys Asp Leu Leu Gly Lys Arg Tyr Ile Glu Leu Phe Arg Ser Thr
 395 400 405
 Ala-Ala-Glu-Val-Gln-Gln-Val-Leu-Asn-Arg-Phe-Ser-Ser-Ala-Pro
 410 415 420
 Leu Ile Pro Leu Pro Thr Pro Pro Ile Ile Pro Val Leu Pro Gln
 425 430 435
 Gln Phe Val Pro Pro Thr Asn Val Arg Asp Cys Ile Arg Leu Arg
 440 445 450
 Gly Leu Pro Tyr Ala Ala Thr Ile Glu Asp Ile Leu Asp Phe Leu
 455 460 465
 Gly Glu Phe Ala Thr Asp Ile Arg Thr His Gly Val His Met Val
 470 475 480
 Leu Asn His Gln Gly Arg Pro Ser Gly Asp Ala Phe Ile Gln Met
 485 490 495
 Lys Ser Ala Asp Arg Ala Phe Met Ala Ala Gln Lys Cys His Lys
 500 505 510
 Lys Asn Met Lys Asp Arg Tyr Val Glu Val Phe Gln Cys Ser Ala
 515 520 525
 Glu Glu Met Asn Phe Val Leu Met Gly Gly Thr Leu Asn Arg Asn
 530 535 540
 Gly Leu Ser Pro Pro Pro Cys Lys Leu Pro Cys Leu Ser Pro Pro
 545 550 555
 Ser Tyr Thr Phe Pro Ala Pro Ala Ala Val Ile Pro Thr Glu Ala
 560 565 570
 Ala Ile Tyr Gln Pro Ser Val Ile Leu Asn Pro Arg Ala Leu Gln

	575	580	585
Pro Ser Thr Ala Tyr Tyr Pro Ala Gly		Thr Gln Leu Phe Met Asn	
590	595	600	
Tyr Thr Ala Tyr Tyr Pro Ser Pro Pro	Gly Ser Pro Asn Ser	Leu	
605	610	615	
Gly Tyr Phe Pro Thr Ala Ala Asn Leu	Ser Gly Val Pro Pro	Gln	
620	625	630	
Pro Gly Thr Val Val Arg Met Gln Gly	Leu Ala Tyr Asn Thr	Gly	
635	640	645	
Val Lys Glu Ile Leu Asn Phe Phe Gln	Gly Tyr Gln Tyr Ala	Thr	
650	655	660	
Glu Asp Gly Leu Ile His Thr Asn Asp	Gln Ala Arg Thr Leu	Pro	
665	670	675	
Lys Glu Trp Val Cys Ile			
	680		

<210> 13
<211> 408
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2879070CD1

	<400> 13		
Met Ser Ser Leu Val Glu Thr Phe Val Ser Lys Ala Ser Ala Leu			
1	5	10	15
Gln Arg Gln Gly Arg Ala Gly Arg Val Arg Asp Gly Phe Cys Phe			
20	25	30	
Arg Met Tyr Thr Arg Glu Arg Phe Glu Gly Phe Met Asp Tyr Ser			
35	40	45	
Val Pro Glu Ile Leu Arg Val Pro Leu Glu Glu Leu Cys Leu His			
50	55	60	
Ile Met Lys Cys Asn Leu Gly Ser Pro Glu Asp Phe Leu Ser Lys			
65	70	75	
Ala Leu Asp Pro Pro Gln Leu Gln Val Ile Ser Asn Ala Met Asn			
80	85	90	
Leu Leu Arg Lys Ile Gly Ala Cys Glu Leu Asn Glu Pro Lys Leu			
95	100	105	
Thr Pro Leu Gly Gln His Leu Ala Ala Leu Pro Val Asn Val Lys			
110	115	120	
Ile Gly Lys Met Leu Ile Phe Gly Ala Ile Phe Gly Cys Leu Asp			
125	130	135	
Pro Val Ala Thr Leu Ala Ala Val Met Thr Glu Lys Ser Pro Phe			
140	145	150	
Thr Thr Pro Ile Gly Arg Lys Asp Glu Ala Asp Leu Ala Lys Ser			
155	160	165	
Ala Leu Ala Met Ala Asp Ser Asp His Leu Thr Ile Tyr Asn Ala			
170	175	180	
Tyr Leu Gly Trp Lys Lys Ala Arg Gln Glu Gly Gly Tyr Arg Ser			
185	190	195	
Glu Ile Thr Tyr Cys Arg Arg Asn Phe Leu Asn Arg Thr Ser Leu			
200	205	210	
Leu Thr Leu Glu Asp Val Lys Gln Glu Leu Ile Lys Leu Val Lys			
215	220	225	

WO 00/15799

Ala Ala Gly Phe Ser Ser Ser Thr Thr Ser Thr Ser Trp Glu Gly
 230 235 240
 Asn Arg Ala Ser Gln Thr Leu Ser Phe Gln Glu Ile Ala Leu Leu
 245 250 255
 Lys Ala Val Leu Val Ala Gly Leu Tyr Asp Asn Val Gly Lys Ile
 260 265 270
 Ile Tyr Thr Lys Ser Val Asp Val Thr Glu Lys Leu Ala Cys Ile
 275 280 285
 Val Glu Thr Ala Gln Gly Lys Ala Gln Val His Pro Ser Ser Val
 290 295 300
 Asn Arg Asp Leu Gln Thr His Gly Trp Leu Leu Tyr Gln Glu Lys
 305 310 315
 Ile Arg Tyr Ala Arg Val Tyr Leu Arg Glu Thr Thr Leu Ile Thr
 320 325 330
 Pro Phe Pro Val Leu Leu Phe Gly Gly Asp Ile Glu Val Gln His
 335 340 345
 Arg Glu Arg Leu Leu Ser Ile Asp Gly Trp Ile Tyr Phe Gln Ala
 350 355 360
 Pro Val Lys Ile Ala Val Ile Phe Lys Gln Leu Arg Val Leu Ile
 365 370 375
 Asp Ser Val Leu Arg Lys Lys Leu Glu Asn Pro Lys Met Ser Leu
 380 385 390
 Glu Asn Asp Lys Ile Leu Gln Ile Ile Thr Glu Leu Ile Lys Thr
 395 400 405
 Glu Asn Asn

<210> 14
 <211> 351
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3093845CD1

<400> 14
 Met Ile Pro Lys Ser Tyr Thr Glu Glu Asp Leu Arg Glu Lys Phe
 1 5 10 15
 Lys Val Tyr Gly Asp Ile Glu Tyr Cys Ser Ile Ile Lys Asn Lys
 20 25 30
 Val Thr Gly Glu Ser Lys Gly Leu Gly Tyr Val Arg Tyr Leu Lys
 35 40 45
 Pro Ser Gln Ala Ala Gln Ala Ile Glu Asn Cys Asp Arg Ser Phe
 50 55 60
 Arg Ala Ile Leu Ala Glu Pro Lys Asn Lys Ala Ser Glu Ser Ser
 65 70 75
 Glu Gln Asp Tyr Tyr Ser Asn Met Arg Gln Glu Ala Leu Gly His
 80 85 90
 Glu Pro Arg Val Asn Met Phe Pro Phe Val Gly Glu Gln Gln Ser
 95 100 105
 Glu Phe Ser Ser Phe Asp Lys Asn Asp Ser Arg Gly Gln Glu Ala
 110 115 120
 Ile Ser Lys Arg Leu Ser Val Val Ser Arg Val Pro Phe Thr Glu
 125 130 135
 Glu Gln Leu Phe Ser Ile Phe Asp Ile Val Pro Gly Leu Glu Tyr

140	145	150
Cys Glu Val Gln Arg Asp Pro Tyr Ser Asn Tyr Gly His Gly Val		
155	160	165
Val Gln Tyr Phe Asn Val Ala Ser Ala Ile Tyr Ala Lys Tyr Lys		
170	175	180
Leu His Gly Phe Gln Tyr Pro Pro Gly Asn Arg Ile Gly Val Ser		
185	190	195
Phe Ile Asp Asp Gly Ser Asn Ala Thr Asp Leu Leu Arg Lys Met		
200	205	210
Ala Thr Gln Met Val Ala Ala Gln Leu Ala Ser Met Val Trp Asn		
215	220	225
Asn Pro Ser Gln Gln Gln Phe Met Gln Phe Gly Gly Ser Ser Gly		
230	235	240
Ser Gln Leu Pro Gln Ile Gln Thr Asp Val Val Leu Pro Ser Cys		
245	250	255
Lys Lys Lys Ala Pro Ala Glu Thr Pro Val Lys Glu Arg Leu Phe		
260	265	270
Ile Val Phe Asn Pro His Pro Leu Pro Leu Asp Val Leu Glu Asp		
275	280	285
Ile Phe Cys Arg Phe Gly Asn Leu Ile Glu Val Tyr Leu Val Ser		
290	295	300
Gly Lys Asn Val Gly Tyr Ala Lys Tyr Ala Asp Arg Ile Ser Ala		
305	310	315
Asn Asp Ala Ile Ala Thr Leu His Gly Lys Ile Leu Asn Gly Val		
320	325	330
Arg Leu Lys Val Met Leu Ala Asp Ser Pro Arg Glu Glu Ser Asn		
335	340	345
Lys Arg Gln Arg Thr Tyr		
350		

<210> 15
<211> 472
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3685685CD1

1	5	10	15
Met Gly Gln Ser Gly Arg Ser Arg His Gln Lys Arg Ala Arg Ala			
20	25	30	
Gln Ala Gln Leu Arg Asn Leu Glu Ala Tyr Ala Ala Asn Pro His			
35	40	45	
Ser Phe Val Phe Thr Arg Gly Cys Thr Gly Arg Asn Ile Arg Gln			
50	55	60	
Leu Ser Leu Asp Val Arg Arg Val Met Glu Pro Leu Thr Ala Ser			
65	70	75	
Arg Leu Gln Val Arg Lys Lys Asn Ser Leu Lys Asp Cys Val Ala			
80	85	90	
Val Ala Gly Pro Leu Gly Val Thr His Phe Leu Ile Leu Ser Lys			
95	100	105	
Thr Glu Thr Asn Val Tyr Phe Lys Leu Met Arg Leu Pro Gly Gly			
110	115	120	
Pro Thr Leu Thr Phe Gln Val Lys Lys Tyr Ser Leu Val Arg Asp			

WO 00/15799

Val	Val	Ser	Ser	Leu	Arg	Arg	His	Arg	Met	His	Glu	Gln	Gln	Phe
									125	130				135
Ala	His	Pro	Pro	Leu	Leu	Val	Leu	Asn	Ser	Phe	Gly	Pro	His	Gly
									140	145				150
Met	His	Val	Lys	Leu	Met	Ala	Thr	Met	Phe	Gln	Asn	Leu	Phe	Pro
									155	160				165
Ser	Ile	Asn	Val	His	Lys	Val	Asn	Leu	Asn	Thr	Ile	Lys	Arg	Cys
									170	175				180
Leu	Leu	Ile	Asp	Tyr	Asn	Pro	Asp	Ser	Gln	Glu	Leu	Asp	Phe	Arg
									185	190				195
His	Tyr	Ile	Lys	Val	Val	Pro	Val	Gly	Ala	Ser	Arg	Gly	Met	Lys
									200	205				210
Lys	Leu	Leu	Gln	Glu	Lys	Phe	Pro	Asn	Met	Ser	Arg	Leu	Gln	Asp
									215	220				225
Ile	Ser	Glu	Leu	Leu	Ala	Thr	Gly	Ala	Gly	Leu	Ser	Glu	Ser	Glu
									230	235				240
Ala	Glu	Pro	Asp	Gly	Asp	His	Asn	Ile	Thr	Glu	Leu	Pro	Gln	Ala
									245	250				255
Val	Ala	Gly	Arg	Gly	Asn	Met	Arg	Ala	Gln	Gln	Ser	Ala	Val	Arg
									260	265				270
Leu	Thr	Glu	Ile	Gly	Pro	Arg	Met	Thr	Leu	Gln	Leu	Ile	Lys	Val
									275	280				285
Gln	Glu	Gly	Val	Gly	Glu	Gly	Lys	Val	Met	Phe	His	Ser	Phe	Val
									290	295				300
Ser	Lys	Thr	Glu	Glu	Glu	Leu	Gln	Ala	Ile	Leu	Glu	Ala	Lys	Glu
									305	310				315
Lys	Lys	Leu	Arg	Leu	Lys	Ala	Gln	Arg	Gln	Ala	Gln	Gln	Ala	Gln
									320	325				330
Asn	Val	Gln	Arg	Lys	Gln	Glu	Gln	Arg	Glu	Ala	His	Arg	Lys	Lys
									335	340				345
Ser	Leu	Glu	Gly	Met	Lys	Lys	Ala	Arg	Val	Gly	Gly	Ser	Asp	Glu
									350	355				360
Glu	Ala	Ser	Gly	Ile	Pro	Ser	Arg	Thr	Ala	Ser	Leu	Glu	Leu	Gly
									365	370				375
Glu	Asp	Asp	Asp	Glu	Gln	Glu	Asp	Asp	Asp	Ile	Glu	Tyr	Phe	Cys
									380	385				390
Gln	Ala	Val	Gly	Glu	Ala	Pro	Ser	Glu	Asp	Leu	Phe	Pro	Glu	Ala
									395	400				405
Lys	Gln	Lys	Arg	Leu	Ala	Lys	Ser	Pro	Gly	Arg	Lys	Arg	Lys	Arg
									410	415				420
Trp	Glu	Met	Asp	Arg	Gly	Arg	Gly	Arg	Leu	Cys	Asp	Gln	Lys	Phe
									425	430				435
Pro	Lys	Thr	Lys	Asp	Lys	Ser	Gln	Gly	Ala	Gln	Ala	Arg	Arg	Gly
									440	445				450
Pro	Arg	Gly	Ala	Ser	Arg	Asp	Gly	Gly	Arg	Gly	Arg	Gly	Arg	Gly
									455	460				465
Arg	Pro	Gly	Lys	Arg	Val	Ala								
									470					

<210> 16
<211> 616
<212> PRT
<213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3825977CD1

<400> 16
 Met Ser Ser Leu Ala Val Arg Asp Pro Ala Met Asp Arg Ser Leu
 1 5 10 15
 Arg Ser Val Phe Val Gly Asn Ile Pro Tyr Glu Ala Thr Glu Glu
 20 25 30
 Gln Leu Lys Asp Ile Phe Ser Glu Val Gly Ser Val Val Ser Phe
 35 40 45
 Arg Leu Val Tyr Asp Arg Glu Thr Gly Lys Pro Lys Gly Tyr Gly
 50 55 60
 Phe Cys Glu Tyr Gln Asp Gln Glu Thr Ala Leu Ser Ala Met Arg
 65 70 75
 Asn Leu Asn Gly Arg Glu Phe Ser Gly Arg Ala Leu Arg Val Asp
 80 85 90
 Asn Ala Ala Ser Glu Lys Asn Lys Glu Glu Leu Lys Ser Leu Gly
 95 100 105
 Pro Ala Ala Pro Ile Ile Asp Ser Pro Tyr Gly Asp Pro Ile Asp
 110 115 120
 Pro Glu Asp Ala Pro Glu Ser Ile Thr Arg Ala Val Ala Ser Leu
 125 130 135
 Pro Pro Glu Gln Met Phe Glu Leu Met Lys Gln Met Lys Leu Cys
 140 145 150
 Val Gln Asn Ser His Gln Glu Ala Arg Asn Met Leu Leu Gln Asn
 155 160 165
 Pro Gln Leu Ala Tyr Ala Leu Leu Gln Ala Gln Val Val Met Arg
 170 175 180
 Ile Met Asp Pro Glu Ile Ala Leu Lys Ile Leu His Arg Lys Ile
 185 190 195
 His Val Thr Pro Leu Ile Pro Gly Lys Ser Gln Ser Val Ser Val
 200 205 210
 Ser Gly Pro Gly Pro Gly Pro Gly Pro Gly Leu Cys Pro Gly Pro
 215 220 225
 Asn Val Leu Leu Asn Gln Gln Asn Pro Pro Ala Pro Gln Pro Gln
 230 235 240
 His Leu Ala Arg Arg Pro Val Lys Asp Ile Pro Pro Leu Met Gln
 245 250 255
 Thr Pro Ile Gln Gly Gly Ile Pro Ala Pro Gly Pro Ile Pro Ala
 260 265 270
 Ala Val Pro Gly Ala Gly Pro Gly Ser Leu Thr Pro Gly Gly Ala
 275 280 285
 Met Gln Pro Gln Leu Gly Met Pro Gly Val Gly Pro Val Pro Leu
 290 295 300
 Glu Arg Gly Gln Val Gln Met Ser Asp Pro Arg Ala Pro Ile Pro
 305 310 315
 Arg Gly Pro Val Thr Pro Gly Gly Leu Pro Pro Arg Gly Leu Leu
 320 325 330
 Gly Asp Ala Pro Asn Asp Pro Arg Gly Gly Thr Leu Leu Ser Val
 335 340 345
 Thr Gly Glu Val Glu Pro Arg Gly Tyr Leu Gly Pro Pro His Gln
 350 355 360
 Gly Pro Pro Met His His Ala Ser Gly His Asp Thr Arg Gly Pro
 365 370 375
 Ser Ser His Glu Met Arg Gly Gly Pro Leu Gly Asp Pro Arg Leu
 380 385 390

Leu Ile Gly Glu Pro Arg Gly Pro Met Ile Asp Gln Arg Gly Leu
 395 400 405
 Pro Met Asp Gly Arg Gly Gly Arg Asp Ser Arg Ala Met Glu Thr
 410 415 420
 Arg Ala Met Glu Thr Glu Val Leu Glu Thr Arg Val Met Glu Arg
 425 430 435
 Arg Gly Met Glu Thr Cys Ala Met Glu Thr Arg Gly Met Glu Ala
 440 445 450
 Arg Gly Met Asp Ala Arg Gly Leu Glu Met Arg Gly Pro Val Pro
 455 460 465
 Ser Ser Arg Gly Pro Met Thr Gly Gly Ile Gln Gly Pro Gly Pro
 470 475 480
 Ile Asn Ile Gly Ala Gly Gly Pro Pro Gln Gly Pro Arg Gln Val
 485 490 495
 Pro Gly Ile Ser Gly Val Gly Asn Pro Gly Ala Gly Met Gln Gly
 500 505 510
 Thr Gly Ile Gln Gly Thr Gly Met Gln Gly Ala Gly Ile Gln Gly
 515 520 525
 Gly Gly Met Gln Gly Ala Gly Ile Gln Gly Val Ser Ile Gln Gly
 530 535 540
 Gly Gly Ile Gln Gly Gly Ile Gln Gly Ala Ser Lys Gln Gly
 545 550 555
 Gly Ser Gln Pro Ser Ser Phe Ser Pro Gly Gln Ser Gln Val Thr
 560 565 570
 Pro Gln Asp Gln Glu Lys Ala Ala Leu Ile Met Gln Val Leu Gln
 575 580 585
 Leu Thr Ala Asp Gln Ile Ala Met Leu Pro Pro Glu Gln Arg Gln
 590 595 600
 Ser Ile Leu Ile Leu Lys Glu Gln Ile Gln Lys Ser Thr Gly Ala
 605 610 615
 Ser

<210> 17
 <211> 112
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 4941262CD1

<400> 17
 Met Val Lys Gly Arg Thr Gly Gln Arg Val Arg Leu Tyr Val Arg
 1 5 10 15
 Gly Thr Ile Leu Gly Tyr Lys Arg Ser Lys Ser Asn Gln Tyr Glu
 20 25 30
 Thr Thr Ser Leu Ile Gln Ile Glu Gly Val Asn Thr Lys Glu Asp
 35 40 45
 Val Ala Trp Tyr Ala Gly Lys Arg Met Ala Tyr Ile Tyr Lys Ala
 50 55 60
 Lys Thr Lys Ser Ser Glu Thr Arg Tyr Arg Cys Ile Trp Gly Lys
 65 70 75
 Val Thr Arg Pro His Gly Asn Ser Gly Val Val Arg Ala Lys Phe
 80 85 90
 Lys Ser Asn Leu Pro Pro Glu Ser Met Gly Arg Lys Val Arg Val
 95 100 105

Phe Met Tyr Pro Ser Ser Ile
110

<210> 18
<211> 1872
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No.: 399781CB1

<210> 19
<211> 5897
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No.: 1252206CB1

<400> 19

```

gactcaatat agggaaatttt gccctcgagg caagaattcg gaacgaggaa cctcttgcggg 60
atcgaatctt cactccgct gtctcagcag tctacagcac ggtAACACAA gtggcaagac 120
agccggaaac ccctacccca tccccttatt cagcacatga aataaacaag gggcatccaa 180
atcttgcggc aacgcccccg ggacatgcat cgtcccctgg actctctcaa accccttac 240
cctctggaca gaatgcaggt ccaaccacgc tggatacc tcaaaccct cagacaatga 300
attcacaacc tcaaaccctgt tctccgttt tccagaggcc tcaaatacag cctccttagag 360
ctaccatccc gaacagcagt cttccattc gtccctggtc acagacaccc actgcagtgt 420
accaggctaa tcagcacatc atgatggta accatctgcc catgcgtac ccagtgcggg 480
aggggcctca gtactgtata ccacagtacc gtcatagtgg ccctccttat gttggggccc 540
ccccaaaata tccagttcaa ccaccggggc caggtcctt ttatcctgga ccaggacctg 600
gggacttccc caatgcttat ggaacgcctt tttacccaag tcagccgtg tatcagtca 660
cacctatcat agtgcctacg cagcaacacgc cgccctccagc caagagagag aaaaaacta 720
taagaattcg ggatccaaac cagggaggt aagacataac agaggagatt atgtctggag 780
gtggcagcag aaatcctact ccacccatag gaagacccac gtcccacacctt attcctcctc 840
agctgcccag ccaggtcccc gggcacagcc ctgtggttt tggtactgtg gagagcgctc 900
atcttgcgtc cagcacccct gtcactgcag ctacgcacca gaagcaagag gagaagccaa 960
aaccagatcc agtgttaaag tctccctccc cagtccttag gctagtcctc agtggagaga 1020
agaaaagaaca agaaggccag acatctgaaa ctactgcaat agtattccata gcagagcttc 1080
ctctgcctcc atcacctacc actgtttctt ctgttgctcg aagtacaatt gcagccccc 1140
cctcttctgc tcttagtagc caaccaatat tcaccactgc tatagatgac agatgtga 1200
tctcatcccc aagagaagac acaattccta tacccagcct cacatctgc acagaaaacat 1260
cagaccctt accaacaataat gaaaatgatg atgatataatg caagaaaaccc tggtagttag 1320
cacctaatga tattccactg gtttctagta ctaacctaataat taatgaaata aatggagtt 1380
gcgaaaaatt atcagccacg gagagcattt tgaaaatagt aaaacaggaa gtattgcct 1440
tgactcttga attggagatt ctcgaaaatc ccccagaaga aatgaaaactg gaggtagtatcc 1500
cagctcccat cacccttcc acagttccctt ctttccctcc aactcctccca actcctccag 1560
cttctcctcc tcacactcca gtcattgttc ctgtgtctgc cactactgtt agttctccga 1620
gtgctgccccat cacagtccag agagtccctag aggaggacga gagcataaga acttgcctt 1680
gtgaagatgc aaaagagatt cagaacaaaa tagaggtaga agcagatggg caaacagaaag 1740
agattttgga ttctcaaaac ttaaatccaa gaaggagccc tgtccctcg ccaatagcta 1800
taactgtacc aaagacatgg aagaaaccaa aagatcgac ccgaaccact gaagagatgt 1860
tagagggcaga attggagctt aaagctgaag aggagcttc cattgacaaa gtacttgaat 1920
ctgaacaaga taaaatgacg caggggttcc atcctgaaag agaccctct gaccta 1980
aagtgaaagc tggaaagaa aatggagaag aagctgaccc agtacgtaat ggtgtgaga 2040
gtgtttctga gggtaagga atagatgcta attcaggctc cacagatagt tctggtgatg 2100
gggttacatt tccattttaaa ccagaatctt ggaagcctac tgatactgaa ggtAAAGAc 2160
agtatgacag ggagtttctg ctggacttcc agttcatgccc tgcctgtata caaaaaccag 2220
agggcctgccc tcctatcagt gatgtggttc ttgacaagat caaccaaccc aaattgccaa 2280
tgcgaactct ggatcctcga attttgcctc gaggaccaga ctttacacca gccttgcgt 2340
attttggaaag gcagacaccc ggtggaaagag gcttacccctt gttgaatgtt gggtcacgaa 2400
gatctcaacc tggccaaaga agagaaccca gaaagatcat cacagttctt gtaaaagaag 2460
atgtacaccc taaaaaggca gaaaatgcctt ggaagccaa cccaaaacca gacagccaa 2520
ccgatgatecc cgaaaaacatt aaaacccagg agtccccctttag aaaaaggatcgaa agtacatctaa 2580
ataaaattgac accacagatg ttcaatcaac tggatgacca agtgcaccc cttactgtt 2640
acacagagga gggctgaaa ggagtttattt acctggcttt tgagaaggctt attgatgaa 2700
ccagtttctc tggatgtttac gcaaaatgtt gtcgtatgtt agtaacgcgtt aagatacc 2760
tggcagacaa gcctggtaac acagtgttattt tccggaaagct gctactgaaac cggtgcaga 2820
aggagtttga aaaagataaa gcaatgtatg atgttttgc gaaagacccgaa aagaacttg 2880
aggctgccccat tggatgttttgc gggatgtt gggatgtt tggatgttcc tggatgttcc 2940
acaaaagcccg gcccggatcc attggcaaca tcaagtttat tggatgttcc tttaaactca 3000
aaatgctgac tgaagccatc atgcacatgt gttggatgtt gctgttccaaag aaccatgtt 3060
aagaatccctt ggagtttctg tggatgttcc tcaccacccat tggatgttcc tggatgttcc 3120
aaaaagcaaa gcccacgtatg gaccatgtt ttaatcagat ggagaaaattt gtgaaagaaa 3180
aaaaaaacccctt atctaggatt cggatgttcc tcaccacccat tggatgttcc tggatgttcc 3240
attgggtatc tcgaagagac gatcaaggcc cttaaactat cgaacagatt cacaagagg 3300
ctaaaataga agaacaagaa gagcaagga aggtccaccc actcatgacc aaagagaaga 3360

```

gaagaccagg tgtccagaga gtggacgaag gtgggtggaa cactgtacaa gggccaaga 3420
 acagtccggt actggacccc taaaattcc taaaaatcac taaggctaca attgatgaaa 3480
 aaattcagct ggtacctaaa gcacagctag gcagctgggg aaaaggcagc agtggtgag 3540
 caaaggcaag tgagactgat gccttacggt caagtgcac cagttaaac agattctctg 3600
 cccgtcaacc tccagcaccc tcagggtcca cgccatccac gcctgttagag tttgattccc 3660
 gaaggacctt aactagtctgt ggaagtatgg gcagggagaa gaatgacaag ccccttccat 3720
 ctgcaacagc tcggccaaat acttcatga ggggtggcag cagtaaagac ctgctagaca 3780
 atcagtctca agaagagcag cggagagaga tgctggagac cgtgaagcag ctcacaggag 3840
 gtgtggatgt ggagaggaac agcactgagg ctgagcggaa taaaacaagg gagtcagcaa 3900
 aaccagaaat tttagcaatg tcagctcatg acaaggctgc attatcagaa gaggaaactgg 3960
 agaggaagtc gaaatctatc attgatgaat ttctacacat taatgattt aaggaaagcca 4020
 tgcaagtgtgt ggaagagctg aatgcccagg gcctactaca ttttttggag agagtggag 4080
 tggagtccac cctggaaagg agccagatca ccagggatca catgggcca 4140
 agctggtaca gtcagaaaaa ctcagcaaac aggactttt caaaggttt tcagaaactt 4200
 tggaaattggc agatgacatg gccattgata ttccccatat ttgggtgtac cttgctgaac 4260
 tggtagcccc catgtaaaaa gaaggtggaa tctccatgag agaacttacc atagaattt 4320
 gcaaaccctt acttcctgtt ggaagagctg gggcttgct atctgaaata ttgcacctac 4380
 tatgcaaaca aatgagccat aagaaagtgg gagccttatg gaggaggct gacccagct 4440
 ggaaggactt ttaccagaa ggagaagatg tacataattt ttttttggag cagaagttgg 4500
 acttcataga gtctgacagt ccctgttcc ctgaagcact ttcaaaagaaa gaactgtctg 4560
 ccgaagagct gtataagcga ctcgagaaac tcattattga ggacaaagcg aatgatgaac 4620
 agatcttga ctggtagag gctaattctag acgaaatcca gatgagttca ctttacattcc 4680
 ttagagctt aatgactgct gttgtaaag cagctattat agccgactct tcttacattca 4740
 gagtggacac tgctgttatac aagcagagag tgccgatctt actcaagtac cttagactcag 4800
 atacagagaa ggaactgcaa gcactttatg cactacaagc atcgatagta aaacttgatc 4860
 aacctgccaa ttgtctgcgg atgtttttt attgtctata tgacgaggag gtgatctccg 4920
 agatgcctt ctacaaatgg gagagcagca aggacccctgc agagcagaat gggaaaggcg 4980
 tggctctgaa atctgtcagc gcattcttca cgtggctgcg ggaagcagaa gaggaggctg 5040
 aggataacta aaacttcaaa tacacaaaat gaaacaaaag aaacaatttta agtattttt 5100
 taaaaagttt cacgtcttcg ccaatcacag tgcaagcaagg ccaattctcg cagaaaacccc 5160
 cacgtgtcga cgagtggag aggggaaaga gaaaaaaaaa tgatcatgga gggaaaaagg 5220
 actggataaa agtaaacttc aaacctttagg gcgggagcac taaaacccaa atacatgtat 5280
 tatttataaa aaatatttc ttgttttatac tttttttt aaacaaggac tcatacttta 5340
 aaaaatgttt agcaaaaaaa aaaaaagttt agaacttttta atttattttt aggactgcaa 5400
 atgccagtgt aatttttaa ttgcagttt ctgttaacaa cttgtataat agaaaaggcag 5460
 agaaataaat ttccctcccc ttcaagatgc acctcatgtt ttgttttaagg tatagcattt 5520
 agtccagatt tgagaaaagtt tgggttgaaac aaggttaagaa agatttttt ttttttggca 5580
 tcaaatctt ctgcctgcct ctcagcttgc ttcaaaaaat taaaaaaatc acaatagtaa 5640
 tcaaaacata cataacattg aaacagaagg aaatgtgtg gaccacagaa ctccaagaat 5700
 tggtaaaaaa aaaaaaagtg ctacccttgag aaaagtactc ttaataactct tggaaatctt 5760
 agagcaactt taaggcttgtt aaatacatag aacaaatatt taaaaaaaca aaaagaaatt 5820
 gactcagttac tatttctttt cactttgaaa atataaagaa caaaataaag acaaacattg 5880
 caagttaaaa aaaaaaaa 5897

<210> 20
 <211> 919
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2950994CB1

<400> 20
 tgggtgggg gcccacccc agagctcaga gcctcctgac ttctgtgccc tctgtcccag 60
 gtggagatgt ttgggtgac agccctggg ctggagcaga gctcacagct gctggaggag 120

ttcctgtccc ttcagatgga gatcttgaca gagctggct tgcaactccg ggtcctggat 180
 atgcccaccc aagaactggg cctcccccgc taccgcaagt ttgacattga ggcctggatg 240
 ccaggcccgag gccgcttgg agaggtcacc agtgcattca actgcacaga ctccagagc 300
 cgccgcctcc acatcatgtt ccagaccgag gctggggagc tgcagttgc ccacacggtg 360
 aacgccaccc cctgtgtgtt ccccccgcctt ctatcgcc tcctggagag taaccagcag 420
 aaggacggct cagtgttgtt gccccctgcc ctcagtcctt acctcggcac tgatcgatc 480
 acagccccata cccacgtgcc tctccagttt acggccccca accagcccccg gaagcctggg 540
 ctgcctggcc agectgtgtt aagctaagaa cccacccaca gcagccctcg ggggtgtcac 600
 tgcttcctgg agttcaggag accccggaca cctggggacct gtgttgcgtga gcccgtcctg 660
 acatctgtgtt tcttcctgtc agtccacgc ccggggccct ggaccacggg gtccacctct 720
 cctctgtctt tgctgcctca gagtcagtca ctgaccctgt tatcatttag ggtcccaagt 780
 ggaaggcagga cgtctgggct ttacggttt agggacagga gaagcagagg aagaggcttc 840
 catccctctt ctccttccatgt gctgagcaaa aagtccccaa taaatggtca 900
 ggacaaaagaa aaaaaaaaaa 919

<210> 21
 <211> 1867
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3461657CB1

<400> 21
 cggggcacca cagccatgtg ctgcgttagcg tcaggcgcta ccggccggccg gggcgctgtg 60
 gagaatgagg aggacctgccc agaactgtcg gacagcgggg acgaggccgc ctggggagat 120
 gaggacgatg cagatctccc ccacggcaag cagcagacc cctgcctgtt ctgttaacagg 180
 ttattcacat ctgctgaaga aacattttca cactgttaagt ctgagcatca gtttatattt 240
 gacagcatgg ttcataaaca tggacttggaa ttttatggat acattaagct aataaatttt 300
 attagactta agaattctac agttgagtttccatgttatacaaccc agtgccttg 360
 gagaaaagaag agtatttgaa gccagtatttta gaagatgacc ttttacttca atttgatgt 420
 gaagatctt atgaaccgggt gtcagtttccatgttatacc ccaatggact cagtgaaaat 480
 acatctgttg ttgaaaaattt gaaacatatg gaagccaggg cactgtctgc tgaagccgca 540
 ttggccagag cacgtgagga tctgcaaaaa atgaaacaaat ttgctcagga ttttgcgtatg 600
 cacacagatg tcagaacctg ctcgtcatct actagtgtca ttgcggaccc ccaggaggat 660
 gaggatgggt ttttatttcag ctcatacggg cattatggga tacatgaaga aatgctaaag 720
 gacaaaatac gaacagaaag ctaccggat ttcatatacc aaaatccaca tatcttcaaa 780
 gacaaggtag tttggatgt tgggtgtgga actggaaattt tctctatgtt tgctgctaaa 840
 gctggggcga agaagggttct tggagtgtat caatctgaaa tactttacca ggcataatggat 900
 attataagac taaataaact tgaagataact attacactaa ttaaaggaaa gattgaagaa 960
 gttcatctt ctgttagaaaa agtagatgtt atcatatctg agtggatggg ctatttttt 1020
 ctgtttgagt ctatgttaga ttctgtcctt tatgcaaaga acaaataactt ggc当地agga 1080
 ggctcggtct accctgacat ttgcactatc acccttggat cagtgtgtga tgtgaataaaa 1140
 catgctgata gaattgtttt ttggggatgtat gtcctatggct tcaagatgtc ctgc当地gaag 1200
 aaagcagttt ttccagaagc tgggtgtgaa gttttagatc cgaagactct tatttcagaa 1260
 ccttgggtt ttaagcatat agattggccat acgacgtctt tctcagattt ggaattttca 1320
 tcagattttt ccctgaaaaat cacaaggaca tccatgtgca cggcaatttc tggctacttt 1380
 gatatatattt ttgagaagaa ttgccacaac agggtcgtgt tctctacggg ccctcagagc 1440
 accaaaaaacac actggaaaca aacagtattt ctactggaaa aaccattttc agttaaagca 1500
 ggtgaaggct taaaaggaaa ggtcacatgtt cacaagaata agaaagatcc acgttctctc 1560
 accgtgaccc tcacgttgaa taattcaact ccaaactttagt gtctccagtg aaacagccat 1620
 aaaagcacac taccttggatgtt gggggtagag tgggtcagca ggaggagat 1680
 ggtttatgtt gggggtagatg gaccctttcc taatgagcct cctcaataag 1740
 agagaagtcc tcattgtggg aatctgacat agttcagctg aggaagagaa tcagctgatc 1800

ctcatggctt gccacgtaat cattttctta gacgtttgct ccaccagatt taaccaaatg 1860
taactcc 1867

<210> 22
<211> 702
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 053076CB1

<400> 22
gaagttagtg atcgaaagca tggcgtcggt ggtgttggcg ctgaggaccc ggacagccgt 60
tacatccttg ctaagccccca ctccggctac agctcttgct gtcagatacg catccaagaa 120
gtcggttgggt agctccaaaa acctcggtgg aaagtcatca ggcagacgcc aaggcattaa 180
gaaaatggaa ggtcaactatg ttcatgctgg gaacatcatt gcaacacacg gccatttccg 240
ctggcacccca ggtgccatg tgggtgttgg gaagaataaa tgtctgtatg cccttggaaaga 300
ggggatagtc cgctacacta aggaggctta cgtgcctcat cccagaaaaca cggaggctgt 360
ggatctgatc accaggctgc ccaagggtgc tgcgtctac aagacttttgc tccacgtgg 420
tcctgccaag cctgaggggca ctttcaactt ggtagctatg ctttgcgttc ctgttggggc 480
catcgacag agactggagc ccaggtgaca ggagatgggtg ataccagaag tcaagggttg 540
gggtggcgac acggccccc gaggaagagg tctgcttgc ggtgactctg caggagactc 600
tgaagtgact gctggaaac ctttggggag acctgacctg gggccaaaaaa taaagtgagc 660
cagcgtcatg aacgcatgct atttagggac aaaaaaaaaaa aa 702

<210> 23
<211> 2459
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1292379CB1

<400> 23
tgcaaaagaa acaaagcaca ttggcttagta gcaataaaaa aagtaaagtc aatacagcat 60
tgaggaattt aaggtatcgt cggggcattt acaaatgcatt tgagtgttgc tccgaaataa 120
aagattttgc aaaccacttt cctacgtacg tccactgttag ttttgcaga tacaacacta 180
gctgttagcaa agcctatgtt aatcatatgtt tgagcttca tagtaaccgt ccaagcaaaa 240
ggttttgtat tttaagaag cattcagaaa atctccgggg cattactcta gtgtgcctta 300
atttgattt cctaagtgtat gtttctggct tagataatat ggctacacac ttaagtcaac 360
ataaaactca tacttgc当地 gttgtatgc agaaatttc tggttgcatac ccaacttctg 420
agcaccttca tgaattaaaaaa aaagaagctc ccgcaagga acaagaacct gtgtctaagg 480
aaattgcaag acctaacatg gctgaaagag aaacagaaaac atcaaattct gaaagtaaac 540
aagataaaagc tgcttcatttca aaagaaaaaa atggatgtaa tgcaatttca tttgaaggct 600
catcaacaac aaaaagtggaa gaaagcataa cagtttcaga taaggaaaaat gaaacctgtc 660
ttgcagacca gggaaactggc tcaaaaaaaca tcgtcagttg tgattcaaat attgggtcag 720
ataaaagtggaa aaagaaaaaa caaaatacaac acgtttgtca gggaaatggag ttgaagatgt 780
gccaaagtgc agaaaaacata atcttatctg atcagattaa agatcacaac tccagtgtc 840
ccagatttca ttccaaagaat attaaggatt tgcgatttagc atcagataat gtaagcatttgc 900
atcagttttt gagaaaaaga catgaacctg aatctgttag ttctgtatgtt agcgagcaag 960
gcagtattca tttggAACCT ctgactccat ccgaggtact tgagtatgaa gcccacagaga 1020
ttcttcagaa aggtgttgggt gatccttcag ccaagactga tgaagtatgt tctgtatcaaa 1080
cagatgacat tccttggagga aataacccta gcacaacaga ggcaacagta gacccgttggaaag 1140

atgaaaaaaga aagaagttga aattagtcat ttaagtttc agtgtaccaa cgataagggc 1200
atttggaca gtgttatcag gtgagtcag tggctgtt gtagttcag aaatggaaat 1260
atgttaaggga ggtcacacat acactttacc tgtatgttca acctatgtta tcaaacaac 1320
caattccacca ataatacgat gatttagtg gattcccaa aagttttaa aaacacgaaac 1380
aggattttaa tgataattaa atttgcagtg gaaaggctc atttaatggt ttcaagaa 1440
atgggatttg gttgctgaca tgaattgtg atattgtaa tatttataaa gcctttcaaa 1500
cttccatcaa tcctaagcta aaaatcttta ttacctgtat atcctttca gttaactgag 1560
aggaaggat ttggaaacca tgtactttt gggagtaatt gattaaaac aatggctgat 1620
tggcattgtt aatgaaggct ttatttgta ggatgtatgt gtaaatggg gcatgcttag 1680
agtactaaat tgatctaattg agaatttggg tgaacataaa cttatattt gatttaatat 1740
aacattccag tcagacgcat gtaaacagaa tatttgaatc tttgtacctc catacaagtg 1800
ttagcctgcc aggctgtaa cttaccttaa taaaacttc agtggaaagtg gaattattaa 1860
gatataaatt tatatttgtt cttttgtca gtgtgtaa gttgttagaaa ttctttgatg 1920
tattagttgt attaatgtaa agtagaaacc cattgttggaa actctgttag ctattatgt 1980
ttaaatattg tttaatgtat cttccctaga aataggccca taaaaatggt ctggaaagcca 2040
aaccaaagta tggtataatg tagatattgt aaagcgttaa actggaaaaca tgcctggca 2100
tgtattcagc catgtttaaag tgactttct gtaattgtaa aataaaaact tcaaatggg 2160
cctaaaacag tgatgtaaaa gaactggttt tggaaattta gcctaatttata tctataagat 2220
ggctgtcaaa ttgatTTTC agttctttt atcatctaga atataataga tatagaaatg 2280
aataatatga agaacagtag tttgcTTTGA aataactaata aacttttatt taaaatgctt 2340
catTTTact tctttaaatg tgcttggat tctttaaattt tggttcactg aatgttcaat 2400
gttttaatg gcgattaaaa tactctgttatatatgtat ttttgagta aatatttcc 2459

<210> 24
<211> 1015
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No.: 1437783CB1

```

<400> 24
gaaggggact cgcgcagat ccgtgtact tgcgtccgct acagtatgtc aatcgcttgc 60
cccagcacag ctgtcattct tctctacaga agagcttctc ctcataact gggatgatt 120
acagttcttc ctaaaaaaagg atggctgctc ttttctaaa gaggttaaca ctacaaactg 180
taaagtctga aaatagttgc attagatgtt ttggtaaaca catcctgcaa aagacagcac 240
cagcacagtt gtccccattt gcttctgccc caagactctc cttcctaatt catgcaaaag 300
ccttagtac cgctgaagac acccagaatg aaggaaaaaa gacaaaaaaag aataaaacag 360
cttttagtaa cgttggaaaga aaaatagtc agcgagttat tcacttattt gatgagaagg 420
gcaatgattt gggaaacatg caccgagcaa atgtgattag acttatggat gagcgagacc 480
tgcgacttgtt tcaaaggaac accagcacag aacctgcaga gtatcagctc atgacaggat 540
tgcagatctt ccaggagcgg cagagctga gggagatgga gaaggcgaac cccaaaactg 600
gaccaaccct gagaaggaa ctgatttgtt cttcaaatat tggacaacat gattggaca 660
caaagactaa acagattcag cagtggatta agaaaaaaaca cctagtccag attaccataa 720
agaaaaggaaa aaatgttagac gtgtcagaaa atgaaatgga ggagatattt catcaaatac 780
tccagactat gcctggataa gctacattct catctaggcc acaagctgtt caaggagggaa 840
aagcttaat gtgttctt cgtgcttga gcaaaaatga ggagaaggca tataaaagaaa 900
ctcaagagac ccaggaaaga gacacttga acaaagacca tgaaaatgat aaggaatcaa 960
atqttctqca tcagtaattt taataaagaa aagcatgctc tgagaaaaaa aaaaa 1015

```

<210> 25
<211> 2211
<212> DNA
<213> *Homo sapiens*

<220>
 <221> misc_feature
 <223> Incyte ID No.: 1557635CB1

<400> 25
 taaacattac gagattggct tggattctgt cggatggact tggggctagc tgccggcgaaa 60
 ctggaggagg ccagataacc atgtcagcca cagttgtaga tgcagttaat gctgcacccc 120
 tatacggttc caaagaaatg agtttggaaag aaccaaagaa gatgaccaga gaggactgga 180
 gaaagaagaa ggagcttagaa gaacagcga aattggcaa tgctcctgca gaagttgatg 240
 aagaaggaaa agacatcaac ccccatattc ctcagttat ttcttcagtg ccatggtata 300
 ttgatccttc aaaaagacct acettaaaac accagagacc acaaccagaa aaacaaaagc 360
 agttcagctc atctggagaa tggtaacaaga ggggtgtaaa agagaattcc ataattacta 420
 agtaccgcaa aggacatgt gaaaattgtg gggccatgac acacaaaaag aaagactgct 480
 ttgagagacc taggcgagtt ggagccaaat ttacaggtac taatatacg ccagatgaac 540
 atgtccagcc tcaactgatg tttgactatg atggaaagag ggatcggtgg aatggctaca 600
 atccagaaga acacatgaaa attgttgaag agtatgccaa agttgattt gcaaaaacgaa 660
 cattgaaagc ccagaaactc caagaggaat tagcctcagg aaaattagtg gaacaggcta 720
 attctccaaa acaccagtgg ggagaagagg aaccaaattc tcagacggaa aaagatcata 780
 atagtgaaga tgaggatgaa gataaatatg cagatgatgat tgacatgcct ggacagaatt 840
 ttgactccaa gagacgaatt actgtccgga atctcaggat tcgagaagat attgcaaaat 900
 atttgaggaa ttttagatcca aattctgcct actatgatcc aaaaactaga gcaatgagag 960
 agaatccta tgccaatgca gggaaagaaatc cagatgaagt gagttatgt ggagataact 1020
 ttgttaggta cacaggagat accatttcaa tggctcagac acagttgtt gcatggaaag 1080
 cctatgacaa gggatctgaa gtgcattcac aggcagatcc tacaaggacta gagctgtt 1140
 ataagtccctt caaagtcaaa aaagaagatt tcaaaagaaca gcagaaagaa agcatcctgg 1200
 aaaagtatgg tggccaaagaa cattttggatg cccctccagc tgaattgtt tttagccaga 1260
 ctgaagacta tggggagttc tcaagacatg ggacagtcataaaggagacag gagcggctg 1320
 ttgcctgctc taagtatgag gaggatgtga agatccacaa tcacacacat atctggggat 1380
 cgtactggaa agaaggccga tggggataca aatgtgtca ctctttttc aagtattcct 1440
 attgtactgg agaagctggg aaggagattt ttaactctga ggagtgtatt ataaatgaga 1500
 taactggggaa agaatctgtg aaaaaaccttc aaacctcat ggagctgcat caagaaaaac 1560
 tgaaagagga aaagaagaag aaaaaaaga aaaagaagaa gcatcgaaag agcagttcg 1620
 atagtgtatgaa tgaagaaaaag aagcatgaaa aattgaaaaa ggcactgaaac gcagaggagg 1680
 cccgccttct tcatgtcaag gagaccatgc agattgtatg gaggaaagcgg ccttacaata 1740
 gcatgtatgaa aactcgagaa cctactgaa agggaaatgga ggcataataga atgaaaacgct 1800
 agaggccaga tgacccatg gcctctttcc ttggacagta gcaactagtc agaagaccat 1860
 ccaagataga tgcagctgat acattttttt cagttttta ttgtatgtt tagatagaaa 1920
 aatcccttgtt tattttttt gctgcctggc tttataaaat atttcagatg cctcacagta 1980
 agttcactcc tttccatact gaggaaacaa gaaaagaaga agaggcacat gaagtgtgct 2040
 ttggggataa gaattttaaa ttggattaag attttttttt cagttttttt tatttttttta 2100
 tttttttttt tgagacggag tcttgctctg tcgcccagge tgaagtgcgg tggcgcgatc 2160
 tcggctcaact gcaagctcca cctcccaggt tcacgccatt ctccctgcctc a 2211

<210> 26
 <211> 1446
 <212> DNA
 <213> Homo sapiens

<220>
 <221>
 <222> 1437
 <223> a or g or c or t, unknown, or other

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2049352CB1

<400> 26
ttgccttggg tcctgccca cagagcggcc tgtctttatc agaggtccct ctgccagggg 60
gaggccccca gagaaaacca gaaagagggt gagagactga ggaagataaa gcgtcccagg 120
gcctcctaca ccagcgccctg agcaggaagc gggagggggcc atgactacga gccctggg 180
ggtaactta gggagggctg tcctaaaacc agaagcttgg agcagaaagt gaaaccctgg 240
tgctccagac aaagatctta gtcgggacta gcggccaag gatgaagcct cacttcagaa 300
acacagtgg a gcaatgtat cgagacacat ttcctacaa ctttataat agacccatcc 360
tttctcgta gaataccgtc tggctgtgt acgaagtgaa aacaagggt ccctaaggc 420
cccctttggg cgcaaagatc ttccgaggcc aggtgtattc cgaacttaag taccaccag 480
agatgagatt cttccactgg ttccagaat ggaggaagct gcatcggtac caggagtatg 540
aggtcacctg gtacatatacc tggagccct gcacaaagtg tacaagggat atggccacgt 600
tcctggccga ggacccgaag gttaccctga ccattttgt tgccgcctc tactacttct 660
gggacccaga ttaccaggag gcgcttcgca gcctgtgtca gaaaagagac ggtccgcgtg 720
ccaccatgaa gatcatgaat tatgacaaat ttccggactg ttggagcaag ttcgtgtaca 780
gccaaagaga gctatttgag ctttggaaata atctgcctaa atattatata ttaactgcaca 840
tcatgctggg ggagattctc agacactcga tggatccacc cacattcaat ttcaacttta 900
acaatgaacc ttgggtcaga ggacggcatg agacttaccc ttgttatgag gtggagcgca 960
tgcacaatga cacctgggtc ctgctgaacc agcgccagggt ctttctatgc aaccaggctc 1020
cacataaaca cggtttcctt gaaggccgc atgcagagct gtgcttcctg gacgtgattc 1080
cctttggaa gctggacctg gaccaggact acagggttac ctgcttcacc tcctggagcc 1140
cctgcttcag ctgtggccag gaaatggcta aattcatttc aaaaaacaaa cacgtgagcc 1200
tgtgcattt cactgcccgc atctatgatg atcaagggat atgtcaggag gggctgcgc 1260
ccctggccga ggctggggcc aaaatttcaa tactgacata cagtgaattt aagcactgct 1320
gggacacccct tggaccac cagggatgtc cttccagcc ctggatgga ctagaggagc 1380
acagccaaacg cctgagtggg aggctgcggg gcattctgca gaatcaggga agctgangga 1440
tgggcc 1446

<210> 27

<211> 1349

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 2231663CB1

<400> 27

cacggaggac taacagtaac accgcccacgc cggcagcaaa gtcattttg gccccccccc 60
cgttccctttt ttcctttta actccttccc tctttgcgaa ttctagaacg gaaccttttt 120
ttaattttcc ccagtagaaaa cgttagaaca atttcgtgaa cgcaatccgg agtgcaccaac 180
atggccggccgg ccgtaaagggtg catgggtaga gccttgatac atcatcaaag gcatagcctt 240
tccaagatgg ttatcagac atcactttgt tcttgttctg taaacatccg agtgcaccaac 300
agacattttg ctgctgctac aaagtcgtca aaaaaacaaa aaaaagggtgc taaagaaaaa 360
acaccagatg agaaaaaaaaga taaaaatgaa aaaataaaag catatcccta tatggaaaggc 420
gaacctgagg atgatgtcta tttaaacgc ttataccgc gacagatata tgaggtggag 480
aaagctgttc acttacttaa gaaatttcaa attcttgact ttactagtcc aaagcaaaatg 540
gttatcttg atttgacact ggatatggca ctgggaaaga agaaaaacgt ggagccattt 600
accagtgttc ttatggcc ataccattt gctccgaaa tcaataaaatg tgctgtatcc 660
acagagaatg catcagaggt caaaatagcg gaagaaaaatg gagctgcatt tgcaggagc 720
actagtctga tacagaagat ttggatgt gaaattgttgc cagacttttgc tggatgttt 780
ccagaaaataa tgcctgaact taatcgatta aggaagaaaac tgaataaaaa atatccaaag 840
ctttctcgaa attccattgg ccgtgacatc cccaaaatgc ttgaatttt taaaatgg 900
catgaaatta agtagatgt gaaaaggag aactttctcc agacaaaaat agcaacattg 960
gatatgtcaa gtgaccagat agctgcaat ctgcaagcag ttatataatgt agtttgtt 1020
cacagaccgc tgaatttggg tcccttgcgt gtcgtgcatt tcctcgtag ttcaacaatgt 1080
gaaggatccattt tactgaagat tgatccattt tggcttaaag aagtaaaaaa tgaagaaatg 1140

gaaaaagaag atgcctaaat gtggtaattt gtaaaaattac tttcagtgg ttaagaagca 1200
 atggagaaaa tgataataca gcataatttt tacatttgat gtctgttat tgatcatact 1260
 tgaaagtggaa cttaacatt gaaaaatcg acagtcattt caagaataag aaaataaaat 1320
 ttctctttg tctgaaaaaa aaaaaaaaaa 1349

<210> 28
 <211> 2596
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2604449CB1

<400> 28
 gtggcctgag gagctcagtt ccctcagcgc ccgtagctc ggcggagtct gcgcgatggg 60
 cgaccggaa aggccggaag cggccggct ggatcaggat gagagatcat cttcagacac 120
 caacgaaagt gaaataaaagt caaatgaaga gccactccta agaaagagtt ctcggcggtt 180
 tgtcatctt ccaatccagt accctgatat ttggaaaatg tataaacagg cacaggctc 240
 cttctggaca gcagaagagg tcgacttatac aaaggatctc ctcactgga acaagctaa 300
 agcagatgag aagtacttca tctctcacat cttagcctt tttgcagcgtt gtgatggat 360
 tgtaaatgaa aatttggtgg agcgctttag tcaggagggtt caggttccag aggctcgctg 420
 tttctatggc tttcaaattt tcacatcgagaa tgttcactca gagatgtaca gtttgcgtat 480
 agacacttac atcagagatc ccaagaaaaag ggaattttta tttaatgc当地 ttgaaaccat 540
 gcccstatgtt aagaaaaaaag cagattggc cttgcgatgg atagcagata gaaaatctac 600
 ttttgggaa agagtgggtt ccttgctgc tggtagaaatgg gtttcttctt caggatctt 660
 tgctgctata ttctggctaa agaagagagg tcttatgcca ggactcactt tttccaatga 720
 actcatcagc agagatgaaag gacttcactg tgactttgtt tgcctgtatgt tccaataactt 780
 agtaaataag ctttcagaag aaagggtcag ggagatcattt gttgatgctg tcaaaattga 840
 gcaggagttt ttaacagaag ccttgcagt tggcctcattt ggaatgaattt gcatatttgc 900
 gaaacagtac attgagttt tagctgacag attacttgc gaaatggat tctcaaaggat 960
 ttttcaggca gaaaatcctt ttgattttat gggaaacattt tctttagaaatgg gaaaacacaaa 1020
 tttcttttaggaa aacgagttt cagatgtca gcttttgca gttatggcag aaaccacacaa 1080
 taacgtcttc accttggatg cagattttta aaaaacctct cgtttttttt ctctataaac 1140
 ttgtcattgg taatgacaga tgccttaatg tgaagcttat ttataatagc aataaaccta 1200
 actggatttg gatgaagaag tcttaataact gacataactgg atttttaatg cactggtttgc 1260
 ttatttggta ttctatcttctt tttccaggc cttccagggtt cacatttttattt tattatgttc 1320
 aataactttgg ttcttagttc ttaaagaatc aagaagttgt gtaatctttt aaaaatatttt 1380
 tcttcagat aaagaaaaaa attaagatgt tgtttacaac tgtttctctt tttttacagt 1440
 acatgtattt aaatcattgc tataataaaat ttaagttcat taggaatataaaaacttgca 1500
 gttctatgtat agattgcattt tattttttttt gtttcattttt atcacaataga aatatggcca 1560
 ggaaggactt gagaagacag tttgatccat tgctttttaga caggactggg ttttgcgttc 1620
 caattatata caataatagt ttttcttaca actaagctgg ccccgccctt gtcttgatat 1680
 taatacatga aattttttata attgtctcat tgctcttattt agaaacatcc atattttctt 1740
 gcttttcttca ttgccatattt ttattttgtc atgaatttgcattt tattttttttt atgttagcgt 1800
 ttgcattttt aaaaatattt cattttgcattt ttacattta aatatgc当地 catcaactgtc 1860
 atagaattcc caaatttcat ttgttagatac tgaactaagg gctaattgtca ggagctgatt 1920
 tttaatgata aagctgcaga tgggctaaat aaaagccaaa ttaatcctac aatcaggtat 1980
 tatgttttta aaccaagttt agtgaattttt tagtggactt gggaaatctt ccccgccaga 2040
 atctggatga atggcacacaa attgaaatctt ctttgcatttcc caccatttcc cttaagtgc 2100
 tctgctcattt tgtaaaaaatg taaagatttg aaagagaatc tcatattccc gaggcattag 2160
 gaagaaagga tttaatccctt tcaattttgg gcttaatctt gttttttttt atgttaagtga 2220
 agatggaaagg ctggagagaaa tgattgtttt ttgtacatgtt aaataaggtc acaatattctt 2280
 tacatactttt gtttacaac tgggttttca ttttttcaaa tggctggccca tttagcaaag 2340
 ttattttacta ttactgtgtt acatagaaatgg ctttatttttgcattt tttttttttt atgttaagtga 2400
 tttgctgaaa tacattatgg tcaatcaagc caagcctgca tggatcagaat ttgtttttttt 2460

ttcaaataaa tttagttgttt tcttattttt ttggcttagt atgttgaat aaactatgg 2520
 atcttcatca ttttgtacat ttccctttt aggaagggtt ctttataagt gcaagggtac 2580
 cctaataaa gaattt 2596

<210> 29
 <211> 2882
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 2604993CB1

<400> 29
 ccctcgagcg gctcgaggc ttttgcata gacgccccggg caactgaata caaaaaggc 60
 aggccctctg cgccctcctt cccacccccc ttccctgcctt gggcggtggag cgccgaccag 120
 gtgtgggctt ggggtgggtgg ttaccgcctt ttgcactagc agtagcaagg aaggggggtg 180
 ggcgctctt cttttctct tagaagaggg tttagcacag gtttttcgt tctcaacttcc 240
 acaccacctt accgcctccc gaccccccct ctccccctcc ccacctatcg tcatgacggc 300
 ctctccggat tacttggtg tgcttttgg gatcaactgt ggggccaccc gggccaagct 360
 aggctcgat gagaaggagt tgatccgt gttctggaaa gtcgtggatc tggccaacaa 420
 gaaggtggga cagttgcacg aagtgttagt tagaccggat cagttggAAC tgacggagga 480
 ctgcaaagaa gaaactaaaa tagacgtcga aacgcctgtcc tcggcgtcgc agctggacca 540
 agccctccga cagtttaacc agtcaactgt caatgaactg aatattggag tagggacttc 600
 cttctgtctc tgtactgtat ggcagttca tgtcaggccaa atccctgcata ctgaggcttc 660
 caagaagaat gtactattac ctgaatgtctt ctattcctt tttgatcttc gaaaagaatt 720
 caagaaatgt tgccctgggt cacctgatata tgacaaaactg gacgttgcacaa atatgacaca 780
 gtatTTaaat tttgagaaga gtatgtcaat ctctcgatata ggagcccttc aagttgaaga 840
 tatgggaat ataatttttag caatgatttc agagccttat aatcacaggt ttccagatcc 900
 agagagagtg aattacaagt ttgaaaagtgg aacttgcagc aagatggaa ttattgtga 960
 taacaccgtt gtcagggcac gaggttacc atggcagttc tcagatcaag atattgcaag 1020
 attcttcaaa ggactcaata ttgccaaggg aggtgcagca ctttgtctga atgctcagg 1080
 tcgaaggAAC ggagaagctc tggtaggtt tgtaagttag gacaccgag accttagcact 1140
 acagaggcac aaacatcaca tggggacccg gtatatttag gtttacaaag caacagggtga 1200
 agatttctt aaaattgtgt gttgtacttc caatgaggtt gcccgatttc tctccaaagga 1260
 aaatcaagtc attgttcgca tgcggggctt cccttcacg gccacagctg aagaagggt 1320
 ggccttcctt ggacagcatt gcccattttc tggggaaag gaaggcatcc tctttgtcacc 1380
 ctacccagat ggtaggccaa cagggacgc ttttgccttc tttgcctgtg aggaatatgc 1440
 acagaatgcg ttgaggaagc ataaaagactt gttggtaaa agatacattt aactcttcag 1500
 gaggcacagca gtcgaagttc agcaggtgtt gatcgatttcc tccctggccc ctctcatcc 1560
 acttccaaacc cttccattt ttcccaactt acctcagccaa tttgtgcctt ctacaaatgt 1620
 tagagactgt atacgccttc gaggttcccttcc ctatgcagcc acaatttggg acatcccttgg 1680
 tttctgggg gagttcgcca cagatattcg tactcatgg gttcacatgg ttttgaatca 1740
 ccaggccgc ccatcaggag atgcctttt ccagatggat tctcgccgaca gaggcattat 1800
 ggctgcacag aagtgtcata aaaaaaacat gaaggacaga tatgttgaag tctttcagtg 1860
 ttcagctgag gagatgaact ttgtgttaat gggggcaact ttaaatcgaa atggcttatac 1920
 cccaccgcca tgaatgttac catgcctgtc tccctcccttcc tacacatttcc cagctccctgc 1980
 tgcagtttattt cctacagaag ctgcatttta ccagccctctt gtgatTTGA atccacgagc 2040
 actgcagcccc tccacagcgt actaccacgc aggcaacttagt ctcttcatttca actacacagc 2100
 gtactatccc agccccccca gttcgcttaa tagtcttgc tactcccttcc cagctgtcaa 2160
 tcttagcggt gtcctccac agcctggcac ggtggtcaga atgcaggggcc tggcctacaa 2220
 tactggagtt aaggaaattt ttaacttctt ccaagggttac cagtagtgcac ccgaggatgg 2280
 acttatacac acaaatgacc aggcaggac tctacccaaa gaatgggtt gtatTTAAGG 2340
 gccccagcag tttagaacatc ctcagaaaag aagtgttgc aagatgtatg gtgatcttga 2400
 aacctccaga cacaagaaaaa cttcttagccaa attcaggggaa agttgtcttca cactcaggct 2460
 gcagtattttt cagcaaactt gattggacaa acggggctgt gccttatctt ttgggtggagt 2520

aaaaaaaaatggagcttagtga agccaaatcg taacttacag caagcagcat gcagcatacc 2580
tggcttttg ctgattgcaa ataggcattt aaaatgtgaa tttggaatca gatgtctcca 2640
ttacttccag taaaagtggc atcataggtg ttcctaagt ttaagtctt ggataaaaac 2700
tccaccatgt tctaccatct ccaccatgaa ctctgttaag gaagcttcat ttttgtatat 2760
tccccgtctt ttcttcttcat ttcccgtgtct tctgcataat catgccttct tgctaagtaa 2820
ttcaagcata agatcttggg ataataaaat cacaatctta ggagaaaagaa taaaatttta 2880
tt 2882

```
<210> 30  
<211> 1777  
<212> DNA  
<213> Homo sapiens
```

<220>
<221> misc_feature
<223> Incyte ID No.: 2879070CB1

<400> 30
gaaatattga aggagcagta ttgatctttt taccaggact tgctcatatt cagcagttgt 60
atgatcttct atcaaatgt agaagatttt attctgaacg atataaaagtg atagctctgc 120
attctattct ttcaacccaa gatcaagctg cagcattcac acttccc&ct ccaggagtca 180
ggaagattgt ttagcaacc aatattgcag agacgggtat cactattct gatgttgat 240
ttgtaattga tactggaga aaaaaagaaa ataagtcacca tgaaagcagt cagatgagtt 300
cttggtgga gacgttgtc agttaagcca gtgcttgca gcgcaggaa agagctggc 360
gggtcagaga tggcttctgt ttccgaatgt acacaagaga aagatttcaa ggctttatgg 420
attattctgt tcctgaaatcc ttacgtgtac ctttggagaa attatgcctt catattatga 480
aatgtaatct tggttctcct gaagatttcc tctccaaagc cttagatcct cctcagctcc 540
aagtgatcag caatgcataa aatttgcctt gaaaaattgg agcttgtgaa taaaatgagc 600
ctaaactgac tccgttgggc caacacccctg cagctttacc tgtgaatgtc aagattggca 660
agatgcttat ttttgggcc atatttggct gccttgaccc agtggcaaca ctagctgcag 720
ttatgacaga gaagtctcct ttaccacac caattggctg aaaagatgaa gcagatctt 780
caaaatcagc tttggccatg gcggattcag accacactgac gatctacaat gcatatcttag 840
gatggaaagaa agcacgacaa gaaggagggtt atcggttctga aatcacatac tgccggagga 900
actttcttaa tagaacatca ctgttaaccc tagaggatgt aaagcaggag ttaataaaagt 960
tggtaaggc agcaggattt tcattttcca caacttctac cagctggaa ggaaacacagag 1020
cctcacagac cctctcattt caagaaatttgc cccttctttaa agctgtactg gtggctggac 1080
tgtatgacaa tgggggaag ataatctata caaagtcaat ggtatgttaca gaaaaattgg 1140
cttgcattgt ggagacggcc caagggaaag cacaagtaca cccatcctca gtaaatcgag 1200
atttgcaaac tcatggatgg ctcttataacc aggagaagat aaggtatgccc agagtgtatt 1260
tgagagaaac taccctaata accccctttc cagtttact ttttgggtt gatataagaag 1320
ttcagcaccc agaacgtctt cttdtcttatttgc atggctggat ctatttcag gcccctgtaa 1380
agatagctgt cattttcaag cagctgagag ttctcatttgc ttcaatggaa agaaaaaaagc 1440
ttgaaaatcc aaagatgtcc ttgaaaatg acaagatttgc cagatcatt acggaatttgc 1500
taaaaacaga gaataactga aactggaaattt catggtcaac tgctttaaaa attaagatga 1560
agatacagtc atgaaattat ctgaaaatgg gtcatcacat taagtatttgc attactttaa 1620
atgttggtac tagccattaa cttaaagggtg gtggggaaaaa agcacataact taaaacatgt 1680
ataattttctt agttccctttt taatgtatgt tattctgttgc gatatttgcctt ctacatttac 1740
aataaattctt ttgggtatttgc aaaaaaaaaaaaaaaa 1777

```
<210> 31  
<211> 1382  
<212> DNA  
<213> Homo sapiens
```

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3093845CB1

<400> 31

tggaaagtac cgagatgtt aagatgaaga acttacaaga atctttgtta tgataccaaa 60
 gtcctacaca gaagaagatc tgccccaaaa atttaagggt tatggagata tcgagtttg 120
 cagcattatt aagaataaag tgactggaga aagtaaagggt ttggctacg tacgatactt 180
 aaaaccatca caagctgcc aagcaataga aaactgtgtat cgaagttta gagcaatctt 240
 ggctgaacct aaaaataaag catctgaatc ctctgaacaa gattattata gtaatatgag 300
 gcaagaagct ttggacatg aaccttaggt aaatatgtt ccatttgtcg gagaacaaca 360
 atctgaattt tcaagttttg acaagaatga tagccgaggc caggaagcaa tctccaaacg 420
 ctgtcagtt gtatcaagag ttcccttcac tgaagaacag ctttcagca ttttgatat 480
 agtaccagga ttgaaatatt gtgaagttca acgagatcct tattcaaatt atggcatgg 540
 agtggttcag tattttatg tagcatcagc tatttatgca aaatacaaat tacatggatt 600
 tcagtagccct cctggaaacc gaataggtgt ttcccttcatt gatgtggaa gtaatgcaac 660
 agatctcctt agaaaaatgg caacacagat ggtagctgc cagctgcac caatgggttg 720
 gaataaccca agtcagcaac aatttatgca atttggagga agctctggat cacagttgcc 780
 tcaaattccag acagatgtt tacttcatc atgcaaaaaaa aaagctcctg ctgaaactcc 840
 tgtgaaagaa agactttta ttgtgtttaa tcctcatcct ttacctttag acgtattaga 900
 agatataattc tgcgttttgc gtaacctgtat cgaagtttac ctgtgtcag gaaaaaatgt 960
 ggggtatgcc aagatgccg atagaataag tgctaattgtat gccattgcca ctctacatgg 1020
 aaagattctg aatgggggtga gacttaaagt tatgctggca gattcgc当地 gagaagaatc 1080
 taacaaacgg caaagaactt actgattctt gagaacaaag actaaataat gacataatcc 1140
 tcagctgact gactgaaaat gtgactggac gatcccctg tggacagttg acagttttt 1200
 tttttttcca tatacctgtat agtctgtgtat cagcattgtt ttgtctggaa agcaggatt 1260
 gctgacatgt attttgaat ccatacatta atgctaaaac gaatataatgta gttgttcctt 1320
 agagcaatat gttgttacgt gtagcagaaaa taaagtttc tttgtttaaa aaaaaaaaaa 1380
 aa 1382

<210> 32
 <211> 1828
 <212> DNA
 <213> Homo_sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3685685CB1

<400> 32

cacggccgtgc cggctctcag gcgcggaaag ttagctgcgc agcgccggaa gcggcggacg 60
 caggaggcct cgtggaggac acagcagcat gggacagtca gggaggtccc ggcaccagaa 120
 gcgcgcgc gcccaggcgc agctcccaa cctcgaggcc tatgcgcga acccgcaactc 180
 gttcgttgc acgcgaggct gcacgggtcg caacatccgg cagctcagcc tggacgtgcg 240
 gccccgtatg gagccgctca ctgcccggc tctgcaggtt cgtaagaaga actcgctgaa 300
 ggactgcgtg gcaatggctg ggcgcctcgg ggtcacacac tttctgtatcc tgagaaaaac 360
 agagaccaat gtctacttta agctgtatgcg cctcccgagga ggcgcacactt tgacccatcca 420
 ggtgaagaag tactcgttgg tgcgtgtatgt ggtctctca ctgcgcggc accgcatgca 480
 cgagcagcag tttccccacc caccctcctt ggtactcaac agcttggcc cccatggat 540
 gcatgtgaag ctatggcca ccatgttcca gaaacctgttc ccctccatca acgtgcacaa 600
 ggtgaacctg aacaccatca agcgctgcct cctcatcgac tacaatcccg actcccgagga 660
 gctggacttc cgccactaca tcaaagggt tctgtgggc gcgagtcgcg ggtatgaagaa 720
 gctgctccag gagaagttcc ccaacatgag ccgcctgcag gacatcagcg agctgctggc 780
 cacggggcgcg gggctgtcgg agagcggaggc agagcctgcac ggcgaccaca acatcacaga 840
 gctgcctcag gctgtcgctg gccgtggcaa catgcgggccc cagcagatg cagtgcggct 900
 caccgagatc ggcccgccgaa tgacactgca gtcatcaag gtccaggagg gcgtcgggaa 960

gggcaaagt atgttccaca gttttgtgag caagacggag gaggagctgc aggccatcct 1020
 ggaagccaag gagaagaagc tgcggctgaa ggcgcagagg caggcccagc aggcccagaa 1080
 tgtcagcgc aagcaggagc agcgggaggc ccacagaaa aagagcctgg agggcatgaa 1140
 gaaggcacgg gtcggggta gtgatgaaga ggcctctgg atcccttcaa ggacggcgag 1200
 cctggagttg ggtgaggacg atgatgaaca ggaagatgat gacatcgagt atttctgcca 1260
 ggcggtggc gaggcgccca gtgaggacct gttccccag gccaagcaga aacggcttgc 1320
 caagtctcca gggcggaaagc ggaagcggtg gaaaatggat cgaggcagg gtcgcctttg 1380
 tgaccagaag tttcccaaga ccaaggacaa gtcggggga gcccgaggca ggcggggcc 1440
 cagaggggct tcccgggatg gtggggcgagg ccggggccgg ggccgcccag ggaagagagt 1500
 ggcctgagcc caagccgcac cggagcagcg gctggattga acgccccaga ttggggcccg 1560
 agatgtggcc ctgggtttcc tttcataaaag gagttgtgtc cccagccctt ccactccagt 1620
 aaagaactga attggccagg ggtccacgtc agcgtttggg atgggggatt ctggagccat 1680
 acaaagcaac ccagagatc ctggggccggc cacaccgag agtccctccc acctgggttc 1740
 ttcctgaaag ctgggtctct ccctaccct gcacggggtt gtttcattg gtggcagcag 1800
 cagccatgag tggccctccc cccagtc 1828

<210> 33
 <211> 2602
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3825977CB1

<400> 33

ccctcgagcg	tgaaatggga	aaggaatcat	gcgactctgt	aataagaagc	tgacagaacc	60
atgaaggatc	agggactgaa	agcattagat	aagttgttag	tggatgcag	gttcacagg	120
caagcggAAC	gattcaagcg	aatttagaaa	gttctgccc	ggatcttgaa	agccgcttcc	180
gttgctcagc	ggaagtgtcg	gtcgcaagag	gacagacgcc	tcgaagaatc	cgctatcgcc	240
tgtctgcaca	accggaatca	tgtcgagttt	ggcggtgaga	gaccggcaaa	tggatcgatc	300
actgcgttcc	gtgttcgtgg	ggaacattcc	atatgaggca	actgaggagc	agttaaagga	360
cattttctcg	gagggtggtt	ctgttgcag	tttccggctg	gtatacgata	gagagacggg	420
aaaacccaag	ggctatggct	tctgcgaata	ccaagaccag	gagaccgcmc	ttagtgccat	480
gcggAACCTC	aatgggcggg	agttcagtgg	gagagcgtt	cggttggaca	atgctgccag	540
tgaaaagaat	aaggaggagt	taaagagcct	tgggcctgca	gcccattt	ttgactcacc	600
ctatggggat	cccatcgatc	cagaagatgc	ccctgaatcg	attaccagag	cagtagccag	660
tctccccccg	gagcagatgt	tttagctgt	gaagcagatg	aagctctgt	tccaaaacag	720
ccaccaggaa	gctcgaaaca	tgttacttca	aaatccacaa	ctggcttatg	cactgttgc	780
ggcacaagta	gtgatgagaa	tcatggatcc	agagattgt	ctgaaaattc	tgcatcgaa	840
gatacatgtc	acaccactga	tcccaggcaa	atctcagtct	gtgtctgtct	ctggccctgg	900
ccctggccct	ggccctgggc	tctgcccagg	acctaatttt	ctgctgaacc	agcagaatcc	960
tccagctct	cagectcagc	atttggctag	aagacctgt	aaggacattc	ctcctctgtat	1020
gcagactct	atccagggtg	gaattccagc	tccaggggca	ataccagctg	cagttcccg	1080
agctggctct	gtttccctaa	ctccctggagg	agcaatgcag	ccccaaactt	gaatgccagg	1140
ggttggccca	gtgcctttag	agcggggaca	agtgcagatg	tcatgcata	gagctcttat	1200
acctcgccga	cccggtactc	ctgggttgtct	gcctctcg	ggactgttag	gagatgctcc	1260
aaatgaccca	cgtggaggga	ctttgccttc	agtcaactgga	gaagtggagc	ccagaggta	1320
tctgggtcca	ccccatcagg	gtccccccat	gcatcatgcc	tctggtcatg	acactcggt	1380
cccttcctca	catgagatga	ggggaggggc	attaggagat	cccagactgc	taattggaga	1440
ccccagagggc	cccatgatag	atcaaagggg	tctacatcg	gatggtagag	gtggtagaga	1500
ttctcgagcg	atggagactc	gtgcattgaa	aactgaggct	ttagagacac	gtgtatgga	1560
gaggagagga	atggagacct	gtgcgtatgaa	aaccagaggg	atggaagcaa	ggggcatgga	1620
tgcaagagga	ttggagatga	ggggccctgt	ccccagttca	agaggcccta	tgactgggt	1680
aattcagggt	cctggtccc	ttaatatagg	ggcagggtggc	cctcctcagg	gaccagaca	1740
gttcccaggc	atttcagggg	tgggaatcc	tggagctggt	atgcagggt	caggcataca	1800

aggaacaggc atgcagggag caggcataca aggaggaggg atgcaggggg caggcataca 1860
 aggagtcagt atacaaggag gaggtatatac agaggaggt atacaggggg caagcaagca 1920
 aggtggaagc cagcttagca gttttagtcc tggcagagc caggtcaactc cacaggatca 1980
 ggagaaggca gcttgatca tgcaggttct tcaactgact gcagatcaga ttgcctatgt 2040
 gccccctgag caaaggcaga gtatcctgat tttaaaggaa caaatccaga aatccactgg 2100
 agcgtcttga aagtttttag aaaatatttgcgtgtacttcaaattttat tctgttagcat 2160
 ggagaatggg tgcaaaaagc tgacttctgt atccccacac ttggattagg gtttccctcc 2220
 tcctagaacc taatcttatt tttgttctt tttttttcc tctgtttcc tttttttttt 2280
 aattgagggg ggggggagga gggagtgcgt ctgttcaattttaactttaaaaataact 2340
 ctgaacatga ttatattatg ccaaataaga ttacaaagaa taagcagcaa tattgaagca 2400
 tctacagtat gttaactaca ttttttaaat gtcgagtaaa acttcgtgaa aactgctcat 2460
 aaagactaaa agttgacgtg taaaacgtt aatgtactaa gatagtttta agattttgg 2520
 ttgtataaca aaataaaaagt ttacccaaaaa gtaaaaaaaaaaaaaaa 2580
 aaaaaaaaaaa aaaaaaaaaaa aa 2602

<210> 34
<211> 566
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4941262CB1

<400> 34
gcagaagtgg aaggaaagcag gcggccacca tggtaaggg acgcacggga cagcgcgtca 60
ggctctacgt ccggggcacc atcctcgat acaagaggc acgtcgaac cagtacgaga 120
ccacgtcgct cattcagatc gagggggta acaccaagga ggacgtcgcg tggtaacgctg 180
gcaagcgcacatc tacaaggcta agaccaagag cagcagacc cgctacagg 240
gcatctgggg caaggtcacc cgcccgacg gcaactcggg cgtcgtccgc gccaaagtca 300
agtccaaacct cccgcctgag tccatggggc gcaaggtaag agtgttcatg tacccgagca 360
gcatctaagg tttttgttgg agtaaaggta gactctaaat ggccatgctt agttcttctc 420
tctgagctt aatgccatg tggtggcac ttagattgtt catgtactga acctgttggaa 480
gttctaccaa aatttgttgcgaaacggctg aacagttgtc ctaatgttat gctataaaca 540
gagcttattt caaaaaaaaaaaaaaa 566

<210> 35
<211> 183
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g2961149

<400> 35
Met Ser Arg Tyr Leu Arg Pro Pro Asn Thr Ser Leu Phe Val Arg
1 5 10 15
Asn Val Ala Asp Asp Thr Arg Ser Glu Asp Leu Arg Arg Glu Phe
20 25 30
Gly Arg Tyr Gly Pro Ile Val Asp Val Tyr Val Pro Leu Asp Phe
35 40 45
Tyr Thr Arg Arg Pro Arg Gly Phe Ala Tyr Val Gln Phe Glu Asp
50 55 60
Val Arg Asp Ala Glu Asp Ala Leu His Asn Leu Asp Arg Lys Trp

65	70	75
Ile Cys Gly Arg Gln Ile Glu Ile Gln	Phe Ala Gln Gly Asp Arg	
80	85	90
Lys Thr Pro Asn Gln Met Lys Ala Lys	Glu Gly Arg Asn Val Tyr	
95	100	105
Ser Ser Ser Arg Tyr Asp Asp Tyr Asp	Arg Tyr Arg Arg Ser Arg	
110	115	120
Ser Arg Ser Tyr Glu Arg Arg Arg Ser	Arg Ser Arg Ser Phe Asp	
125	130	135
Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro	Arg Asn Ser Arg Pro Thr	
140	145	150
Gly Arg Pro Arg Arg Ser Arg Ser His	Ser Asp Asn Asp Arg Pro	
155	160	165
Asn Cys Ser Trp Asn Thr Gln Tyr Ser	Ser Ala Tyr Tyr Thr Ser	
170	175	180
Arg Lys Ile		

<210> 36
<211> 1404
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g2660712

1	5	10	15
Met Ser Gly Ala Arg Thr Ala Ser Thr Pro	Thr Pro Pro Gln Thr		
Gly Gly Gly Leu Glu Pro Gln Ala Asn	Gly Glu Thr Pro Gln Val		
20	25	30	
Ala Val Ile Val Arg Pro Asp Asp Arg	Ser Gln Gly Ala Ile Ile		
35	40	45	
Ala Asp Arg Pro Gly Leu Pro Gly Pro	Glu His Ser Pro Ser Glu		
50	55	60	
Ser Gln Pro Ser Ser Pro Ser Pro Thr	Pro Ser Pro Ser Pro Val		
65	70	75	
Leu Glu Pro Gly Ser Glu Pro Asn Leu	Ala Val Leu Ser Ile Pro		
80	85	90	
Gly Asp Thr Met Thr Thr Ile Gln Met	Ser Val Glu Glu Ser Thr		
95	100	105	
Pro Ile Ser Arg Glu Thr Gly Glu Pro	Tyr Arg Leu Ser Pro Glu		
110	115	120	
Pro Thr Pro Leu Ala Glu Pro Ile Leu	Glu Val Glu Val Thr Leu		
" 125	130	135	
Ser Lys Pro Val Pro Glu Ser Glu Phe	Ser Ser Ser Pro Leu Gln		
140	145	150	
Ala Pro Thr Pro Leu Ala Ser His Thr	Val Glu Ile His Glu Pro		
155	160	165	
Asn Gly Met Val Pro Ser Glu Asp Leu	Glu Pro Glu Val Glu Ser		
170	175	180	
Ser Pro Glu Leu Ala Pro Pro Pro Ala	Cys Pro Ser Glu Ser Pro		
185	190	195	
Val Pro Ile Ala Pro Thr Ala Gln Pro	Glu Glu Leu Leu Asn Gly		
200	205	210	
Ala Pro Ser Pro Pro Ala Val Asp Leu	Ser Pro Val Ser Glu Pro		
215	220	225	

Glu Glu Gln Ala Lys Glu Val Thr Ala Ser Val Ala Pro Pro Thr
 230 235 240
 Ile Pro Ser Ala Thr Pro Ala Thr Ala Pro Ser Ala Thr Ser Pro
 245 250 255
 Ala Gln Glu Glu Glu Met Glu Glu Glu Glu Glu Glu Gly
 260 265 270
 Glu Ala Gly Glu Ala Gly Glu Ala Glu Ser Glu Lys Gly Gly Glu
 275 280 285
 Glu Leu Leu Pro Pro Glu Ser Thr Pro Ile Pro Ala Asn Leu Ser
 290 295 300
 Gln Asn Leu Glu Ala Ala Ala Ala Thr Gln Val Ala Val Ser Val
 305 310 315
 Pro Lys Arg Arg Arg Lys Ile Lys Glu Leu Asn Lys Lys Glu Ala
 320 325 330
 Val Gly Asp Leu Leu Asp Ala Phe Lys Glu Ala Asn Pro Ala Val
 335 340 345
 Pro Glu Val Glu Asn Gln Pro Pro Ala Gly Ser Asn Pro Gly Pro
 350 355 360
 Glu Ser Glu Gly Ser Gly Val Pro Pro Arg Pro Glu Glu Ala Asp
 365 370 375
 Glu Thr Trp Asp Ser Lys Glu Asp Lys Ile His Asn Ala Glu Asn
 380 385 390
 Ile Gln Pro Gly Glu Gln Lys Tyr Glu Tyr Lys Ser Asp Gln Trp
 395 400 405
 Lys Pro Pro Asn Leu Glu Glu Lys Lys Arg Tyr Asp Arg Glu Phe
 410 415 420
 Leu Leu Gly Phe Gln Phe Ile Phe Ala Ser Met Gln Lys Pro Glu
 425 430 435
 Gly Leu Pro His Ile Ser Asp Val Val Leu Asp Lys Ala Asn Lys
 440 445 450
 Thr Pro Leu Arg Pro Leu Asp Pro Thr Arg Leu Gln Gly Ile Asn
 455 460 465
 Cys Gly Pro Asp Phe Thr Pro Ser Phe Ala Asn Leu Gly Arg Thr
 470 475 480
 Thr Leu Ser Thr Arg Gly Pro Pro Arg Gly Gly Pro Gly Gly Glu
 485 490 495
 Leu Pro Arg Gly Pro Gln Ala Gly Leu Gly Pro Arg Arg Ser Gln
 500 505 510
 Gln Gly Pro Arg Lys Glu Pro Arg Lys Ile Ile Ala Thr Val Leu
 515 520 525
 Met Thr Glu Asp Ile Lys Leu Asn Lys Ala Glu Lys Ala Trp Lys
 530 535 540
 Pro Ser Ser Lys Arg Thr Ala Ala Asp Lys Asp Arg Gly Glu Glu
 545 550 555
 Asp Ala Asp Gly Ser Lys Thr Gln Asp Leu Phe Arg Arg Val Arg
 560 565 570
 Ser Ile Leu Asn Lys Leu Thr Pro Gln Met Phe Gln Gln Leu Met
 575 580 585
 Lys Gln Val Thr Gln Leu Ala Ile Asp Thr Glu Glu Arg Leu Lys
 590 595 600
 Gly Val Ile Asp Leu Ile Phe Glu Lys Ala Ile Ser Glu Pro Asn
 605 610 615
 Phe Ser Val Ala Tyr Ala Asn Met Cys Arg Cys Leu Met Ala Leu
 620 625 630
 Lys Val Pro Thr Thr Glu Lys Pro Thr Val Thr Val Asn Phe Arg
 635 640 645
 Lys Leu Leu Leu Asn Arg Cys Gln Lys Glu Phe Glu Lys Asp Lys

650	655	660
Asp Asp Asp Glu Val Phe Glu Lys Lys Gln Lys Glu Met Asp Glu		
665	670	675
Ala Ala Thr Ala Glu Glu Arg Gly Arg Leu Lys Glu Glu Leu Glu		
680	685	690
Glu Ala Arg Asp Ile Ala Arg Arg Arg Ser Leu Gly Asn Ile Lys		
695	700	705
Phe Ile Gly Glu Leu Phe Lys Leu Lys Met Leu Thr Glu Ala Ile		
710	715	720
Met His Asp Cys Val Val Lys Leu Leu Lys Asn His Asp Glu Glu		
725	730	735
Ser Leu Glu Cys Leu Cys Arg Leu Leu Thr Thr Ile Gly Lys Asp		
740	745	750
Leu Asp Phe Glu Lys Ala Lys Pro Arg Met Asp Gln Tyr Phe Asn		
755	760	765
Gln Met Glu Lys Ile Ile Lys Glu Lys Lys Thr Ser Ser Arg Ile		
770	775	780
Arg Phe Met Leu Gln Asp Val Leu Asp Leu Arg Gly Ser Asn Trp		
785	790	795
Val Pro Arg Arg Gly Asp Gln Gly Pro Lys Thr Ile Asp Gln Ile		
800	805	810
His Lys Glu Ala Glu Met Glu Glu His Arg Glu His Ile Lys Val		
815	820	825
Gln Gln Leu Met Ala Lys Gly Ser Asp Lys Arg Arg Gly Gly Pro		
830	835	840
Pro Gly Pro Pro Ile Ser Arg Gly Leu Pro Leu Val Asp Asp Gly		
845	850	855
Gly Trp Asn Thr Val Pro Ile Ser Lys Gly Ser Arg Pro Ile Asp		
860	865	870
Thr Ser Arg Leu Thr Lys Ile Thr Lys Pro Gly Ser Ile Asp Ser		
875	880	885
Asn Asn Gln Leu Phe Ala Pro Gly Gly Arg Leu Ser Trp Gly Lys		
890	895	900
Gly Ser Ser Gly Gly Ser Gly Ala Lys Pro Ser Asp Ala Ala Ser		
905	910	915
Glu Ala Ala Arg Pro Ala Thr Ser Thr Leu Asn Arg Phe Ser Ala		
920	925	930
Leu Gln Gln Ala Val Pro Thr Glu Ser Thr Asp Asn Arg Arg Val		
935	940	945
Val Gln Arg Ser Ser Leu Ser Arg Glu Arg Gly Glu Lys Ala Gly		
950	955	960
Asp Arg Gly Asp Arg Leu Glu Arg Ser Glu Arg Gly Gly Asp Arg		
965	970	975
Gly Asp Arg Leu Asp Arg Ala Arg Thr Pro Ala Thr Lys Arg Ser		
980	985	990
Phe Ser Lys Glu Val Glu Glu Arg Ser Arg Glu Arg Pro Ser Gln		
995	1000	1005
Pro Glu Gly Leu Arg Lys Ala Ala Ser Leu Thr Glu Asp Arg Asp		
1010	1015	1020
Arg Gly Arg Asp Ala Val Lys Arg Glu Ala Ala Leu Pro Pro Val		
1025	1030	1035
Ser Pro Leu Lys Ala Ala Leu Ser Glu Glu Glu Leu Glu Lys Lys		
1040	1045	1050
Ser Lys Ala Ile Ile Glu Glu Tyr Leu His Leu Asn Asp Met Lys		
1055	1060	1065
Glu Ala Val Gln Cys Val Gln Glu Leu Ala Ser Pro Ser Leu Leu		
1070	1075	1080

Phe Ile Phe Val Arg His Gly Val Glu Ser Thr Leu Glu Arg Ser
 1085 1090 1095
 Ala Ile Ala Arg Glu His Met Gly Gln Leu Leu His Gln Leu Leu
 1100 1105 1110
 Cys Ala Gly His Leu Ser Thr Ala Gln Tyr Tyr Gln Gly Leu Tyr
 1115 1120 1125
 Glu Ile Leu Glu Leu Ala Glu Asp Met Glu Ile Asp Ile Pro His
 1130 1135 1140
 Val Trp Leu Tyr Leu Ala Glu Leu Val Thr Pro Ile Leu Gln Glu
 1145 1150 1155
 Gly Gly Val Pro Met Gly Glu Leu Phe Arg Glu Ile Thr Lys Pro
 1160 1165 1170
 Leu Arg Pro Leu Gly Lys Ala Ala Ser Leu Leu Glu Ile Leu
 1175 1180 1185
 Gly Leu Leu Cys Lys Ser Met Gly Pro Lys Lys Val Gly Thr Leu
 1190 1195 1200
 Trp Arg Glu Ala Gly Leu Ser Trp Lys Glu Phe Leu Pro Glu Gly
 1205 1210 1215
 Gln Asp Ile Gly Ala Phe Val Ala Glu Gln Lys Val Glu Tyr Thr
 1220 1225 1230
 Leu Gly Glu Glu Ser Glu Ala Pro Gly Gln Arg Ala Leu Pro Ser
 1235 1240 1245
 Glu Glu Leu Asn Arg Gln Leu Glu Lys Leu Leu Lys Glu Gly Ser
 1250 1255 1260
 Ser Asn Gln Arg Val Phe Asp Trp Ile Glu Ala Asn Leu Ser Glu
 1265 1270 1275
 Gln Gln Ile Val Ser Asn Thr Leu Val Arg Ala Leu Met Thr Ala
 1280 1285 1290
 Val Cys Tyr Ser Ala Ile Ile Phe Glu Thr Pro Leu Arg Val Asp
 1295 1300 1305
 Val Ala Val Leu Lys Ala Arg Ala Lys Leu Leu Gln Lys Tyr Leu
 1310 1315 1320
 Cys Asp Glu Gln Lys Glu Leu Gln Ala Leu Tyr Ala Leu Gln Ala
 1325 1330 1335
 Leu Val Val Thr Leu Glu Gln Pro Pro Asn Leu Leu Arg Met Phe
 1340 1345 1350
 Phe Asp Ala Leu Tyr Asp Glu Asp Val Val Lys Glu Asp Ala Phe
 1355 1360 1365
 Tyr Ser Trp Glu Ser Ser Lys Asp Pro Ala Glu Gln Gln Gly Lys
 1370 1375 1380
 Gly Val Ala Leu Lys Ser Val Thr Ala Phe Phe Lys Trp Leu Arg
 1385 1390 1395
 Glu Ala Glu Glu Glu Ser Asp His Asn
 1400

<210> 37
 <211> 322
 <212> PRT
 <213> Homo sapiens

<300>
 <308> Incyte ID No.: g2440051

<400> 37

Leu Arg Arg Gly Ala Arg Asn Trp Arg Ser Met Ser Thr Gly Glu
 1 5 10 15
 Leu Thr Pro Gln Ser Arg Leu Lys Glu Phe Ser Glu Leu Ala Arg
 20 25 30
 Ala Leu Asn Leu Tyr Arg Met Asp His Leu Gly Asn Tyr Thr Gly
 35 40 45
 His Lys Ser Tyr Tyr Leu Thr Gly Gln Leu Ala Thr Leu Glu Gln
 50 55 60
 Ala Ile Ile Gln Tyr Ala Leu Gln Ala Val Thr Glu His Gly Phe
 65 70 75
 Lys Leu Ile Ser Val Pro Asp Ile Leu Pro Lys Glu Val Ile Glu
 80 85 90
 Ser Cys Gly Met Arg Thr Glu Gly Glu Arg Thr Gln Val Tyr Lys
 95 100 105
 Leu Asp Thr Gly Glu Cys Leu Ser Gly Thr Ser Glu Met Ala Leu
 110 115 120
 Ala Gly Phe Phe Ala Asn Lys Leu Leu Ser Glu Asp Gln Leu Pro
 125 130 135
 Leu Lys Val Thr Ala Val Ser Arg Cys Tyr Arg Ala Glu Thr Ser
 140 145 150
 Gly Leu Gln Glu Glu Lys Gly Ile Tyr Arg Val His Gln Phe Asn
 155 160 165
 Lys Val Glu Met Phe Ala Ile Cys Thr Glu Glu Gln Ser Glu Ala
 170 175 180
 Glu Leu Glu Glu Phe Lys Asn Ile Glu Val Asp Leu Phe Arg Arg
 185 190 195
 Leu Gly Leu Asn Phe Arg Leu Leu Asp Met Pro Pro Cys Glu Leu
 200 205 210
 Gly Ala Pro Ala Tyr Gln Lys Tyr Asp Ile Glu Ala Trp Met Pro
 215 220 225
 Gly Arg Gln Met Trp Gly Glu Ile Ser Ser Cys Ser Asn Cys Thr
 230 235 240
 Asp Tyr Gln Ala Arg Arg Leu Gly Ile Arg Tyr Arg Arg Ser Ala
 245 250 255
 Asp Gly Gln Ile Leu His Ala His Thr Ile Asn Gly Thr Ala Thr
 260 265 270
 Ala Ile Pro Arg Leu Leu Ile Ala Leu Leu Glu Ser Tyr Gln Lys
 275 280 285
 Glu Asp Gly Ile Glu Ile Pro Ala Val Leu Arg Pro Phe Met Asp
 290 295 300
 Asn Gln Glu Leu Ile Thr Arg Asn Lys Arg Ile Pro Glu Thr Lys
 305 310 315
 Leu Val Lys Phe Ile Lys Ala
 320

<210> 38

<211> 343

<212> PRT

<213> Homo sapiens

<300>

<308> Incyte ID No.: g1808648



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :		A3	(11) International Publication Number:	WO 00/15799
C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, A61K 38/17, C12Q 1/68			(43) International Publication Date: 23 March 2000 (23.03.00)	
(21) International Application Number: PCT/US99/21688		Filed on 8 April 1999 (08.04.99)		
(22) International Filing Date: 17 September 1999 (17.09.99)		(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).		
(30) Priority Data: 60/155,246 17 September 1998 (17.09.98) US 09/158,720 22 September 1998 (22.09.98) US Not furnished 22 September 1998 (22.09.98) US 60/069,391 4 November 1998 (04.11.98) US 60/128,660 8 April 1999 (08.04.99) US		(72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, Boulder Creek, CA 95006 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US).		
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications US Not furnished (CIP) Filed on 17 September 1998 (17.09.98) US 09/158,720 (CIP) Filed on 22 September 1998 (22.09.98) US Not furnished (CIP) Filed on 22 September 1998 (22.09.98) US 09/186,815 (CIP) Filed on 4 November 1998 (04.11.98) US 09/156,039 (CIP) Filed on 17 September 1998 (17.09.98) US Not furnished (CIP) Filed on 4 November 1998 (04.11.98) US 60/128,660 (CIP)		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).		
		Published <i>With international search report.</i>		
		(88) Date of publication of the international search report: 13 July 2000 (13.07.00)		

(54) Title: RNA-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human RNA-associated proteins (RNAAP) and polynucleotides which identify and encode RNAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RNAAP.

```

1  [MSRYLRRPPHTSLVVRNVADDTRASEDDLRRE] 399781
1  [MSRYLRRPPHTSLVVRNVADDTRASEDDLRRE] 399781
33 [GRYGFPIVDVYVPLDFTYTRAFPGCFAYVQFED] 399781
33 [GRYGFPIVDVYVPLDFTYTRAFPGCFAYVQFED] 399781
61 [VRDADRALMLIDRKWICGCGTCTGFAAGDD] 399781
61 [VRDADRALMLIDRKWICGCGTCTGFAAGDD] 399781
93 [KTPHOMKAKREGRNVYSSSKRYDDYDRYRRE] 399781
93 [KTPHOMKAKREGRNVYSSSKRYDDYDRYRRE] 399781
121 [SRSYERRRRSRSRPFDTMYTRESYSPRNSRPF] 399781
121 [SRSYERRRRSRSRPFDTMYTRESYSPRNSRPF] 399781
151 [GRPRPRRKEATPTMHDQTAAGIPSTVLLTLD] 399781
151 [GRPRPRRKEATPTMHDQTAAGIPSTVLLTLD] 399781
181 [ELSPFGCRATKREGDFKRPKGGSKVLOVE] 399781
181 [ELSPFGCRATKREGDFKRPKGGSKVLOVE] 399781
209 [GYMFSTTLV] 399781
170 [TYLGS--RRI] 399781

```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

Internat	I Application No
PCT/US 99/21688	

A. CLASSIFICATION OF SUBJECT MATTER	
IPC 7	C12N15/12 C12N5/10 A61K38/17 C12Q1/68
C12N1/21	C07K14/47
	C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMHUM1 [Online] E.M.B.L. Databases Accession Number: AF047448, 17 March 1998 (1998-03-17) YANG L ET AL: "Homo sapiens TLS-associated protein TASR mRNA, complete cds" XP002128498 99.5% identity in 621 bp overlap with SeqIdNo.26 / 100% identity in 155 aa overlap with SeqIdNo.1 abstract -& YANG L ET AL: "Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing" JOURNAL OF BIOLOGICAL CHEMISTRY., vol. 273, no. 43, 23 October 1998 (1998-10-23), pages 27761-27764, XP002128497 the whole document	1-16,19
P,X	-/--	1-16,19

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

24 January 2000

Date of mailing of the international search report

02.05.2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Lonnoy, O

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/US 99/21688

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 679 716 A (MATSUBARA KENICHI ;OKUBO KOUSAKU (JP)) 2 November 1995 (1995-11-02) Human gene signature HUMGS00127 (SeqIdNo.113): 98.9% identity in 349 bp overlap with SeqIdNo.26 -& DATABASE GENESEQ [Online] E.M.B.L. Databases Accession Number: T19113, 4 July 1996 (1996-07-04) MATSUBARA K ET AL: "Human gene signature HUMGS00127" XP002128499 98.9% identity in 349 bp overlap with SeqIdNo.26 abstract ---	3-11
X	US 5 561 222 A (KEENE JACK D ET AL) 1 October 1996 (1996-10-01) figure 6 ---	3-11
A	WO 98 23744 A (INCYTE PHARMA INC ;BANDMAN OLGA (US); GOLI SURYA K (US)) 4 June 1998 (1998-06-04) -----	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 21688

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see additional sheet, subject 1.

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-20 (all partially)

A substantially purified polypeptide of SeqIdNo.1, a fragment of said polypeptide; an isolated and purified polynucleotide encoding said polypeptide, a variant polynucleotide having at least 90% sequence identity to said polynucleotide, a polynucleotide hybridising to said polynucleotide, a polynucleotide having a sequence complementary to said polynucleotide; a detection method using said complementary polynucleotide; a polynucleotide comprising a sequence of SeqIdNo.18 or a fragment thereof, a polynucleotide having at least 90% sequence identity to said polynucleotide of SeqIdNo.18, a polynucleotide complementary to said polynucleotide of SeqIdNo.18; expression vector, host and recombinant method related thereto; a pharmaceutical composition comprising said polypeptide of SeqIdNo.1; an antibody to said polypeptide of SeqIdNo.1 or fragment of said polypeptide of SeqIdNo.1; an agonist to said polypeptide of SeqIdNo.1; an antagonist to said polypeptide of SeqIdNo.1; therapeutic application thereof

Inventions 2-17: claims 1-20 (all partially)

Idem as for subject 1 but limited to each of the polypeptide sequences as in SeqIdNo.2-17 and the corresponding polynucleotide sequences as in SeqIdNo.19-34, wherein respectively invention 2 is limited to SeqIdNo.2 and SeqIdNo.19, invention 3 is limited to SeqIdNo.3 and SeqIdNo.20,..., invention 17 is limited to SeqIdNo.17 and SeqIdNo.34.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box 3.

Although claim 19 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Further defect(s) under Article 17(2)(a):

Continuation of Box 3.

Claims Nos.: 17,18,20

Claims 17, 18 and 20 refer to an agonist, an antagonist and to the use of an antagonist of the polypeptide of claim 1 without giving a true technical characterization. Moreover, no such compound is defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such speculative claims the wording of which is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internal Application No

PCT/US 99/21688

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP 0679716 A	02-11-1995	AU	8116494 A	13-06-1995
		CA	2153480 A	01-06-1995
		WO	9514772 A	01-06-1995
-----	-----	-----	-----	-----
US 5561222 A	01-10-1996	US	5866680 A	02-02-1999
-----	-----	-----	-----	-----
WO 9823744 A	04-06-1998	US	5962226 A	05-10-1999
		AU	7410598 A	22-06-1998
-----	-----	-----	-----	-----